

Indice

Prefazione	1
Capitolo 1 Introduzione.....	3
1.1 <i>Isola di Vulcano</i>	<i>3</i>
1.2 <i>Localizzazione dei pozzi.....</i>	<i>4</i>
Capitolo 2 Dati di riferimento.....	7
2.1 <i>Studio del rapporto Calcio/Solfato</i>	<i>8</i>
2.2 <i>Analisi preliminare</i>	<i>9</i>
Capitolo 3 La distribuzione normale asimmetrica	21
3.1 <i>Cenni preliminari.....</i>	<i>21</i>
3.2 <i>Stima della distribuzione normale asimmetrica</i>	<i>22</i>
3.3 <i>Aspetti salienti della log-verosimiglianza.....</i>	<i>30</i>
3.4 <i>Stima del parametro di asimmetria tramite regressione locale</i>	<i>43</i>
3.5 <i>Stima della mediana tramite regressione locale.....</i>	<i>48</i>
3.6 <i>Stima della mediana tramite verosimiglianza locale.....</i>	<i>51</i>
Capitolo 4 La distribuzione t asimmetrica	57
4.1 <i>Stima della distribuzione t asimmetrica.....</i>	<i>58</i>
4.2 <i>Confronto tra distribuzione normale asimmetrica e t asimmetrica.....</i>	<i>65</i>
4.3 <i>Stima del parametro dei gradi di libertà tramite regressione locale</i>	<i>70</i>
4.4 <i>Stima della mediana tramite regressione locale.....</i>	<i>73</i>
4.5 <i>Stima della mediana tramite verosimiglianza locale.....</i>	<i>76</i>
Capitolo 5 Analisi della variabile in periodi di attività anomala del vulcano	83
5.1 <i>Analisi grafica.....</i>	<i>84</i>
5.2 <i>Stima della mediana tramite verosimiglianza locale.....</i>	<i>89</i>
Conclusioni	93
Bibliografia	95

PREFAZIONE

L'analisi chimica dell'acqua è utile per determinare in modo corretto i vari utilizzi a cui può essere destinata. Noti i cambiamenti ambientali, espressi nella composizione dell'acqua freatica in superficie e in profondità della zona nord occidentale dell'isola di Vulcano nell'arcipelago delle Eolie, si vuole verificare se tali cambiamenti influenzano l'idoneità dell'uso a cui è destinata l'acqua dei pozzi localizzati in quest'area.

A tale scopo, sono stati utilizzati dati forniti dalla dott.ssa Antonella Buccianti, ricercatore presso il Dipartimento di Scienze della Terra dell'Università degli Studi di Firenze, che ringrazio vivamente per il materiale procurato e la disponibilità dimostrata. Tali dati riferiscono ad alcuni campioni d'acqua estratti da ventisei pozzi localizzati nella zona considerata. In particolare viene considerato il rapporto tra due componenti importanti dell'acqua, catione calcio e anione solfato che tendono a combinarsi per formare il solfato di calcio oppure, all'inverso, quando materiali che contengono il solfato di calcio si alterano l'anione ed il catione vengono ceduti all'acqua. Risultano essere componenti importanti dell'acqua, in particolare di quella minerale, e spesso la presenza di solfato è associata a problematiche di inquinamento ambientale.

L'obiettivo è capire se la vicinanza al vulcano influisce o meno sull'andamento dei valori registrati per tale variabile e se è possibile individuare qualche elemento che permette di discriminare i pozzi in base alla loro posizione.

Le analisi vengono svolte tramite il programma *R* che costituisce un linguaggio per il calcolo statistico e la rappresentazione grafica dei dati. È stata utilizzata la versione 2.3.1 ed, in particolare, il pacchetto *sn* 0.4-1.

Capitolo 1

INTRODUZIONE

1.1 Isola di Vulcano

L'arcipelago delle isole Eolie, disposto secondo una stella a tre punte, è localizzato nel Mediterraneo tra la punta estrema della Calabria e la Sicilia, circa 40 km al largo delle coste siciliane ed è formato da sette distinte isole vulcaniche: Alicudi, Filicudi, Panarea, Lipari, Salina, Stromboli e Vulcano. Quest'ultima, terza per dimensioni (22 km²), è la più meridionale, e dista non più di 20 km dalla costa settentrionale della Sicilia. Come tutte le altre isole Eolie, consiste interamente di rocce vulcaniche ed, insieme con Stromboli, è l'unica che presenti ancora un'attività vulcanica.

L'isola di Vulcano ha un'età di circa 125.000 anni e, dal punto di vista geologico-vulcanologico, è una delle isole più recenti dell'arcipelago; il suo vulcanismo ha avuto inizio circa 100.000 anni fa, dando origine agli strato-vulcani della parte meridionale dell'isola. Il primo ciclo di attività si concluse con il collasso della parte più alta dell'edificio e con la conseguente formazione della Caldera del Piano, parzialmente colmata dai prodotti di un'intensa attività che si sviluppava al suo interno. Successivamente si formò la Caldera della Fossa all'interno della quale sorge oggi l'omonimo vulcano ed è l'area con fumarole ed attualmente in uno stato quiescente.

La penisola di Vulcanello apparve inizialmente come nuova isola nel 183 a.C.. Eruzioni recenti si verificarono nel VI e XVI secolo e fu unita a Vulcano per accumulo di sabbia nel 1550.

L'eruzione a La Fossa nel 1888-1890 diede il nome al tipo di attività denominata "vulcaniana" e fu caratterizzata dalla rimozione, dovuta ad esplosione, del tappo craterico e della conseguente emissione di scorie, accompagnata dalla formazione di una scura nube piena di cenere ma senza colate di lava.

Da allora è stata osservata soltanto attività fumarolica dal cono principale. Occasionalmente, nelle fasi di più intensa energia, si verificano fuoriuscite di gas a temperature che raggiungono circa 600-700°C.

Nella zona del Porto di Levane sono presenti emanazioni di gas sulfurei bollenti e fumarole sottomarine.

1.2 Localizzazione dei pozzi

L'isola di Vulcano è un'area vulcanica quiescente dove il coinvolgimento di tipo chimico delle rocce vulcaniche col tempo e l'ingresso dei componenti gassosi derivanti dall'attività fumarolica hanno prodotto cambiamenti ambientali riscontrabili nella composizione dell'acqua freatica in superficie e in profondità. Per comprendere i naturali processi in attività nell'isola è utile analizzare i cambiamenti nella composizione chimica dell'acqua degli strati acquiferi della zona nord occidentale attorno al cratere. Più precisamente, verranno presi in considerazione campioni di acqua raccolti da ventisei pozzi localizzati tra il Porto di Ponente e il Porto di Levante. I dati sono stati forniti dalla dott.ssa Antonella Buccianti del Dipartimento di Scienze della Terra dell'Università degli Studi di Firenze). Nella Figura 1.1 vengono riportate una mappa dell'isola con la localizzazione dei pozzi e un'ulteriore cartina che indica la loro posizione precisa. In quest'ultima è possibile notare una maggiore concentrazione di punti verso la costa del Porto di Levante, mentre vi sono pozzi posizionati in maniera più diradata nella parte sud-ovest.

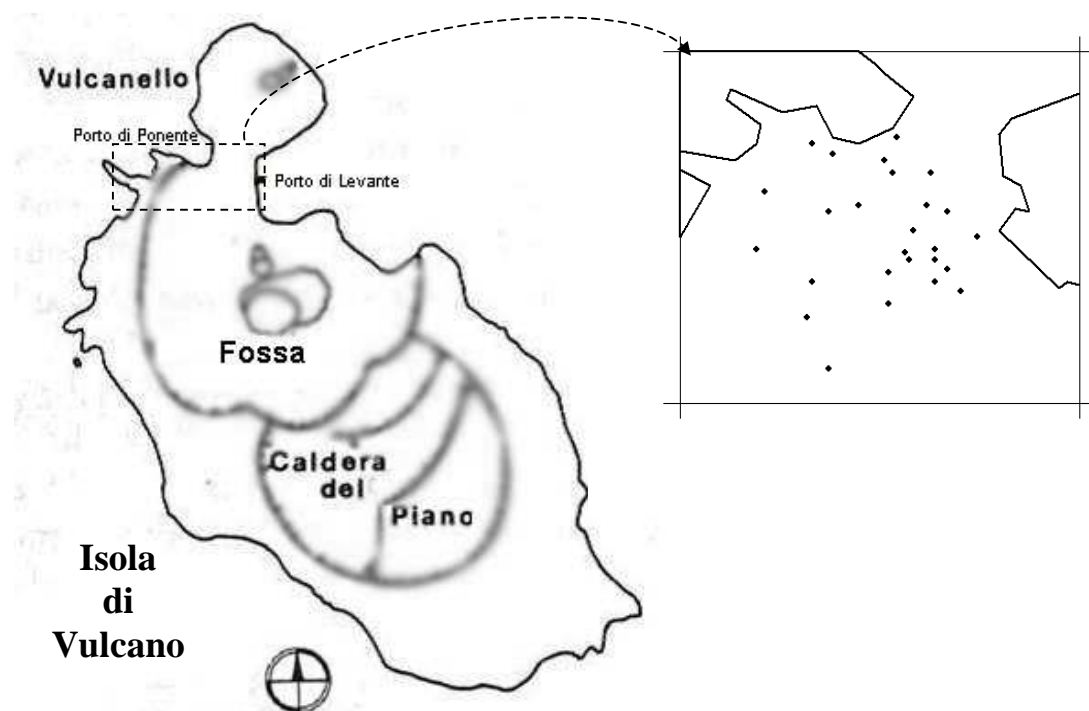


Figura 1.1 Isola di Vulcano (a sinistra) con la localizzazione dei pozzi indicata nel riquadro tratteggiato e (a destra) rappresentazione della posizione precisa, in tale zona, dei pozzi considerati.

Per poter localizzare i ventisei pozzi considerati, si sovrappone un piano cartesiano alla mappa dei pozzi già esistente. Come punto di riferimento, si fissa convenzionalmente l'origine nella punta estrema della costa del Porto di Levante e ogni centimetro, che nella realtà corrisponde a 180 metri, equivale a un'unità. In questo modo è possibile individuare la posizione di ogni pozzo rilevandone le coordinate nel piano cartesiano creato. Ogni pozzo risulta così identificato dalla coppia (x_1, x_2) dove x_1 =asse delle ascisse e x_2 =l'asse delle ordinate. Di seguito vengono riportati il grafico con i pozzi numerati e le relative coordinate .

Nelle analisi che seguono la localizzazione dei pozzi sarà fondamentale e, nonostante non venga sovrapposto ai grafici proposti, si farà sempre riferimento a questo piano cartesiano.

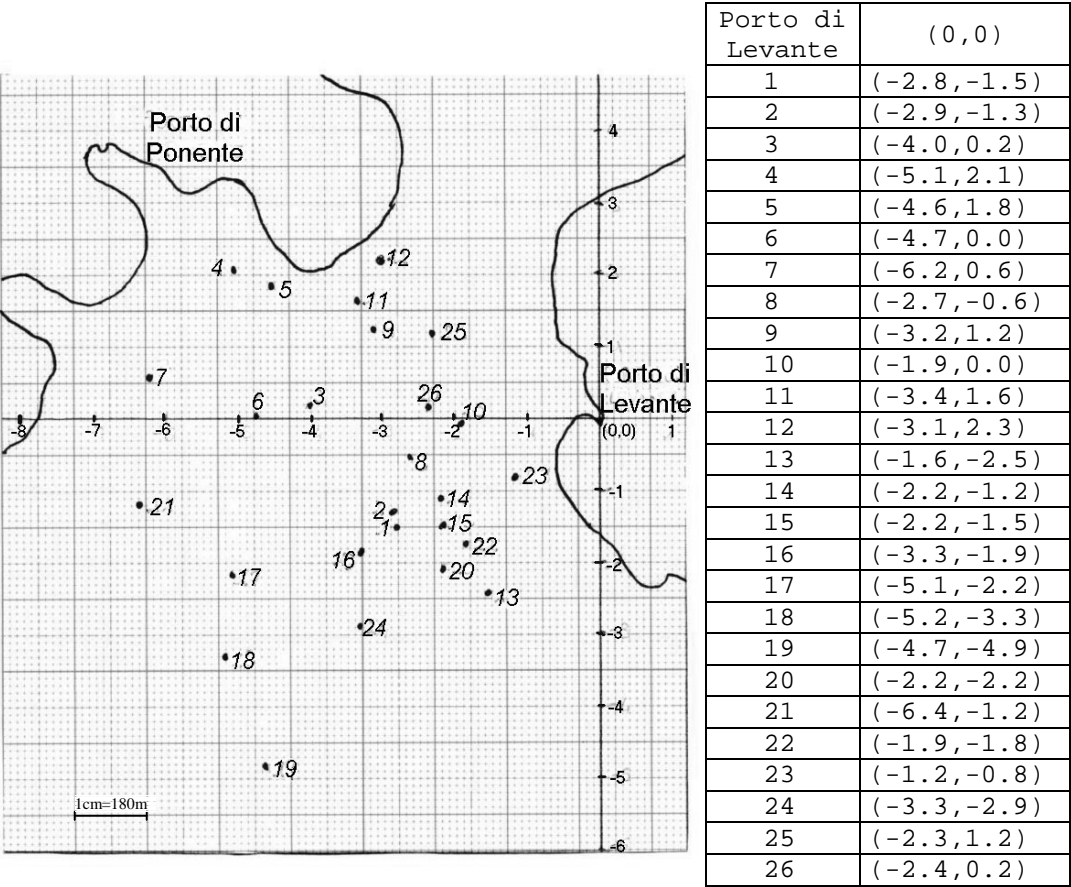


Figura 1.2 Posizione dei pozzi e coordinate che li identificano.

Capitolo 2

DATI DI RIFERIMENTO

L'analisi chimica dell'acqua è utile per determinare in modo corretto i vari utilizzi a cui può essere destinata. Con l'accrescere della conoscenza della chimica dell'acqua, è diventato possibile interpretare i processi evolutivi che ne determinano la composizione e la qualità. Noti i cambiamenti ambientali, espressi nella composizione dell'acqua freatica dei pozzi posti nella zona nord occidentale del cratere attivo dell'isola di Vulcano, si vuole verificare se tali cambiamenti influenzano l'idoneità dell'uso a cui è destinata l'acqua dei pozzi localizzati in quest'area.

A tale scopo, sono stati utilizzati dati forniti dalla dott.ssa Antonella Bucciatti del Dipartimento di Scienze della Terra dell'Università degli Studi di Firenze, che riferiscono ad alcuni campioni d'acqua estratti dai ventisei pozzi considerati (Figura 1.1).

Nelle pagine che seguono, vengono prese in considerazione due variabili scelte in base a criteri geochimici e ciò vuole dire che chimicamente si pensa si comportino in modo analogo, partecipando a simili processi di soluzione/precipitazione. Si considera il rapporto tra due componenti importanti dell'acqua, catione calcio e anione solfato che tendono a combinarsi per formare il solfato di calcio oppure, all'inverso, quando materiali che contengono il solfato di calcio si alterano l'anione ed il catione vengono ceduti all'acqua. Risultano essere componenti importanti dell'acqua, in particolare di quella minerale, e spesso la presenza di solfa-

to è associata a problematiche di inquinamento ambientale.

Il numero di campioni registrati per ogni pozzo è differente, ma l'aspetto di cui si deve tenere maggiormente conto è che tale numerosità campionaria è bassa: per ogni pozzo è a disposizione un numero di osservazioni compreso tra ventisette e trentanove (in totale 977). Per ogni prelievo effettuato è nota anche la data di campionamento (espressa in mese e anno), compresa nell'arco temporale tra giugno 1977 e novembre 1994.

2.1 Studio del rapporto Calcio/Solfato

In questo capitolo si analizza la variabile che considera due componenti importanti dell'acqua, in particolare di quella minerale, ed è utile per analizzare problematiche di inquinamento ambientale.

Tale variabile è espressa tramite il rapporto tra catione Calcio (Ca^{2+}) e anione Solfato (SO_4^{2-}). Le analisi verranno costruite applicando la metodologia per l'analisi di dati composizionali¹ basata sul log-rapporto proposta da Aitchison (1982); particolare importanza di questo approccio è l'estensione dello spazio campionario ristretto a quello reale ampliando, così, la gamma di metodi disponibili per l'analisi della distribuzione dei fenomeni.

La variabile è di tipo quantitativo continuo e, tramite la trasformazione del log-rapporto, assume valori sia positivi che negativi nell'intervallo $[-4.4; 2.6]$.

Non si hanno a disposizione elementi informativi che consentano di formulare un modello statistico di riferimento per rappresentare in maniera semplificata il fenomeno d'interesse quindi si deve innanzitutto procedere con un'esplorazione preliminare dei dati che suggerisca quale sia il modello migliore da adottare.

¹ I dati composizionali sono caratterizzati dal fatto che ogni valore rappresenta una proporzione di un totale prefissato (ad esempio dati espressi in percentuale).

2.2 Analisi preliminare

Come spiegato, si procede tramite un'esplorazione preliminare, utilizzando un approccio non parametrico; in questo modo non si fa riferimento a nessuna formulazione parametrica e non si impone nessuna struttura particolare ai dati, ma si osserva come questi si presentano.

Per analizzare una variabile quantitativa continua (nel caso in esame il log-rapporto tra catione calcio e anione solfato), per ogni pozzo si esegue in prima istanza una sintesi grafica attraverso istogrammi e diagrammi con scatole e baffi. Ad ogni istogramma viene sovrapposta la curva della stima della funzione di densità ignota ottenuta utilizzando uno dei metodi di stima non parametrica più diffuso, il metodo del nucleo.

Di seguito vengono riportati tali rappresentazioni per ogni pozzo dove la variabile y rappresenta il log-rapporto considerato, $y = \ln \left(\frac{Ca^{2+}}{SO_4^{2-}} \right)$.

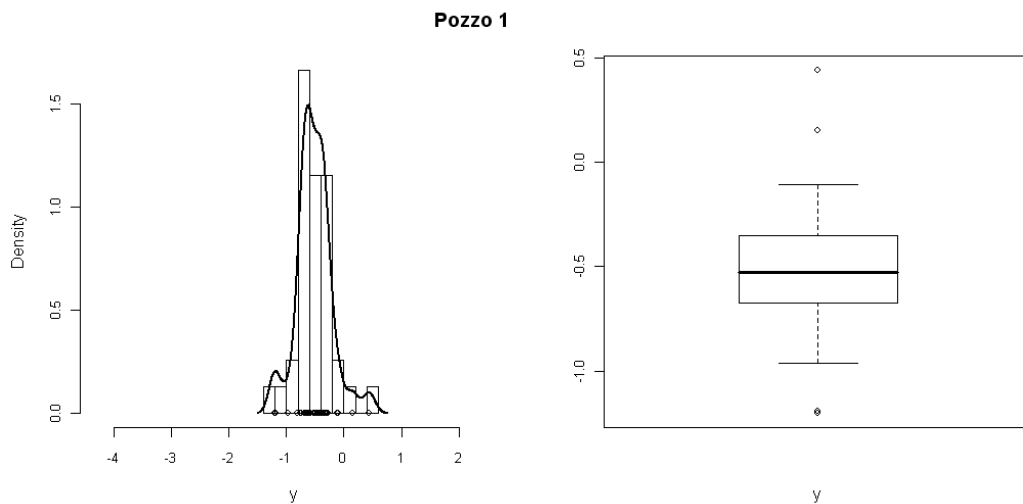


Figura 2.1 Pozzo 1 Istogramma con curva di densità non parametrica stimata (a sinistra), e diagramma con scatola e baffi (a destra).

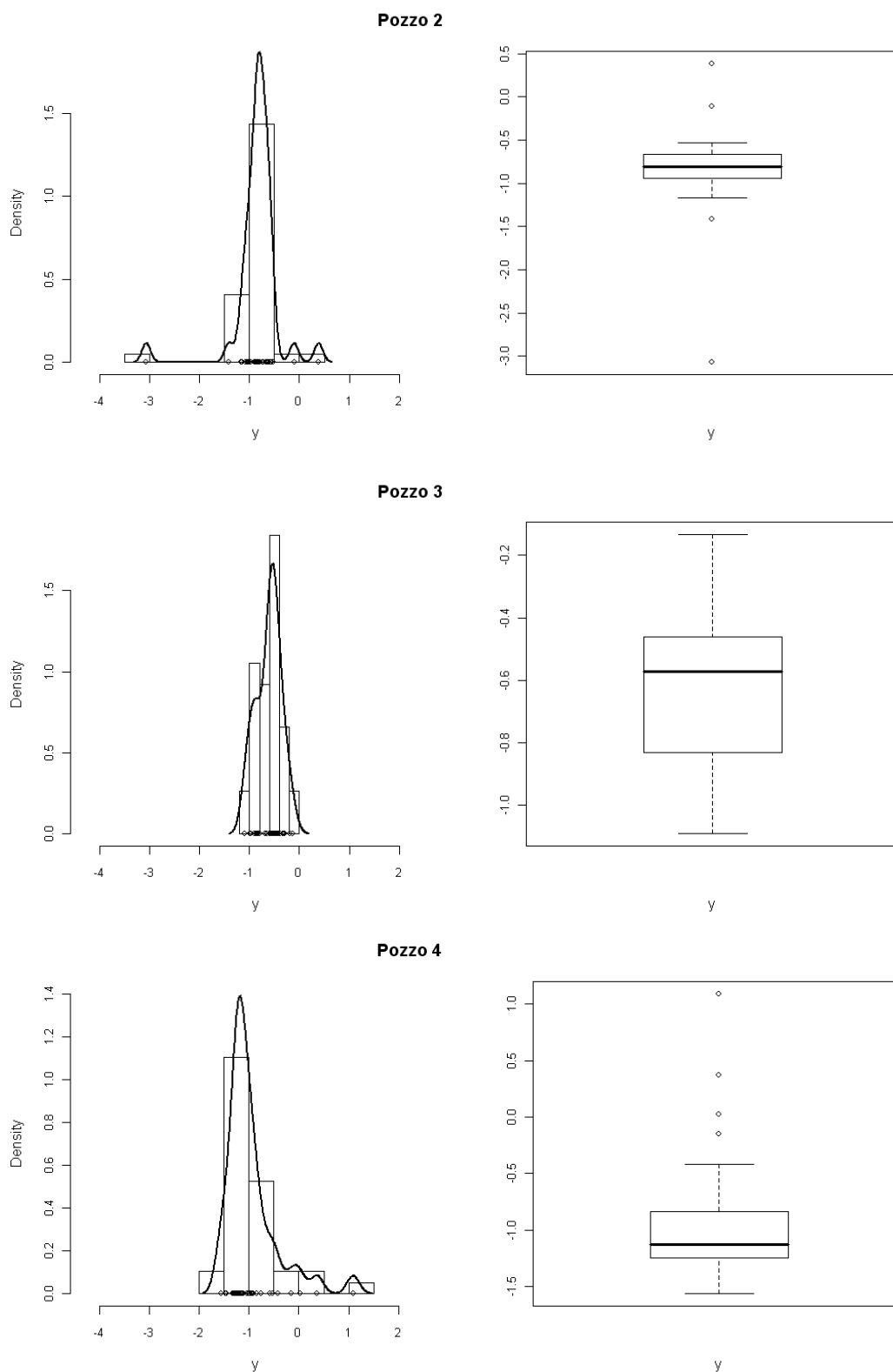
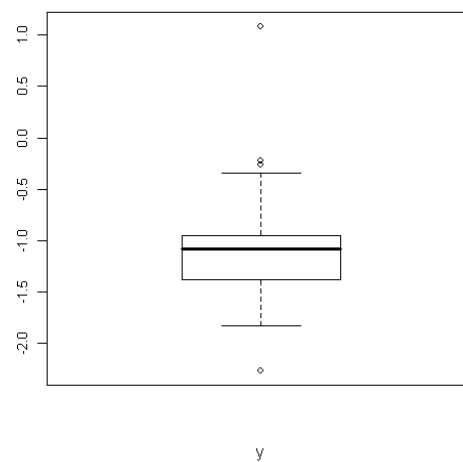
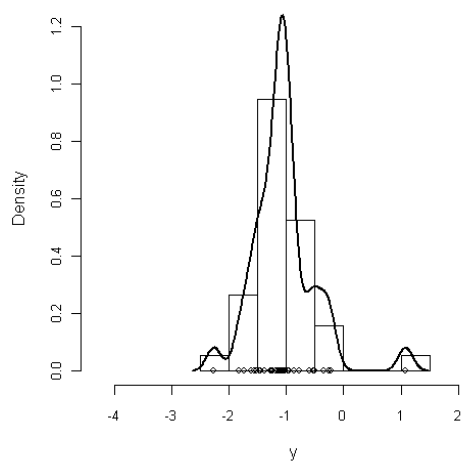
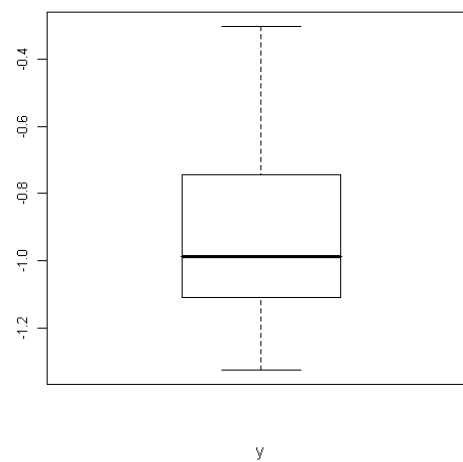
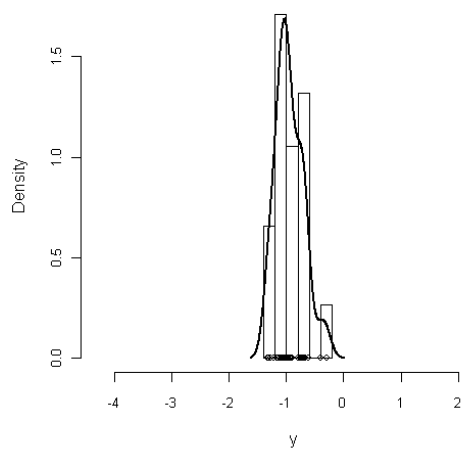


Figura 2.1 Pozzi 2-4 Istogramma con curva di densità non parametrica stimata (a sinistra), e diagramma con scatola e baffi (a destra).

Pozzo 5



Pozzo 6



Pozzo 7

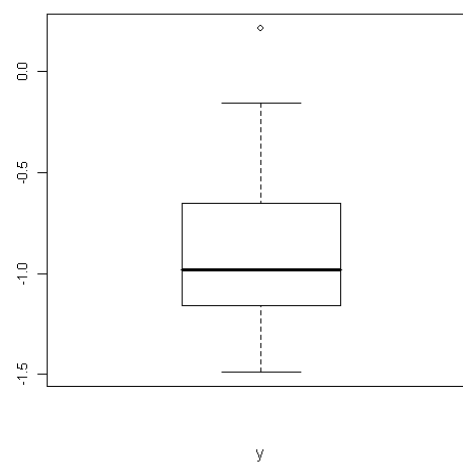
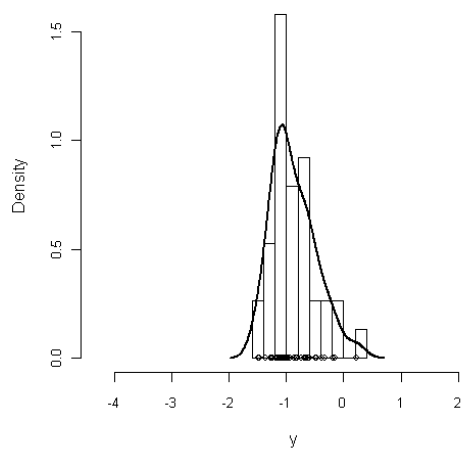
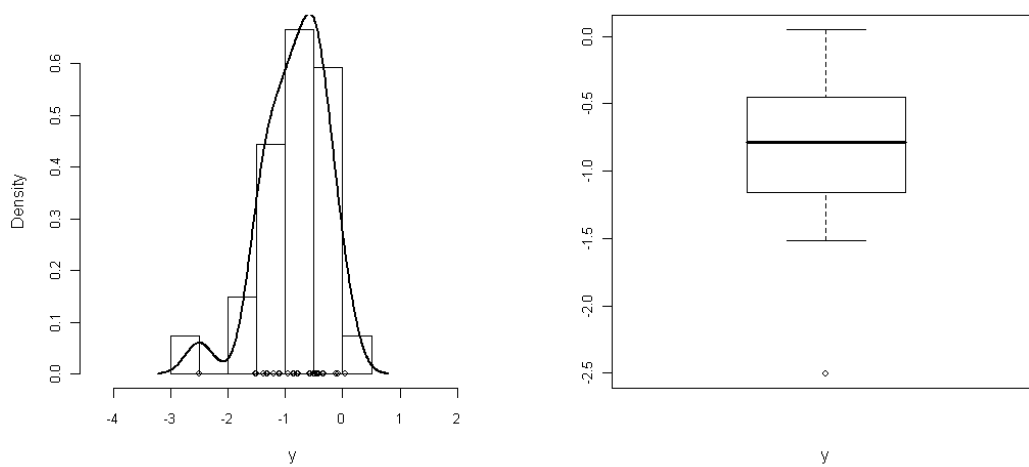
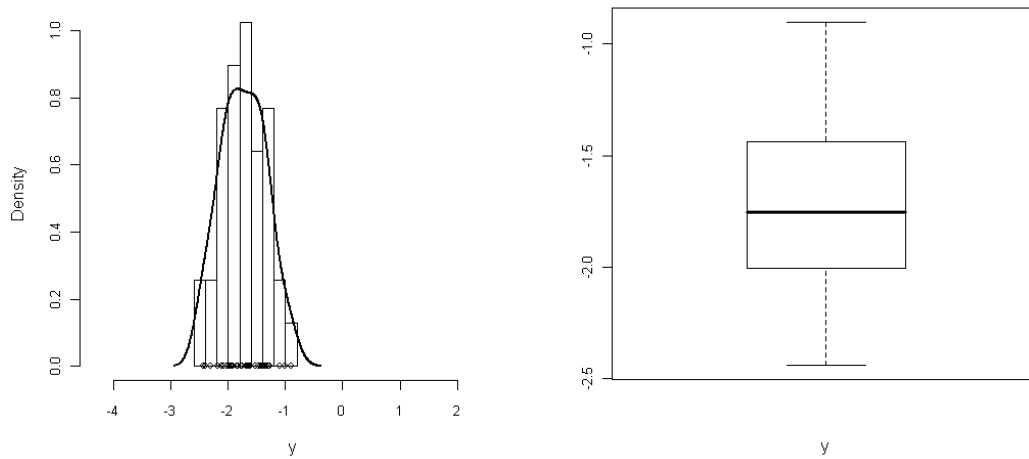


Figura 2.1 Pozzi 5-7 Istogramma con curva di densità non parametrica stimata (a sinistra), e diagramma con scatola e baffi (a destra).

Pozzo 8



Pozzo 9



Pozzo 10

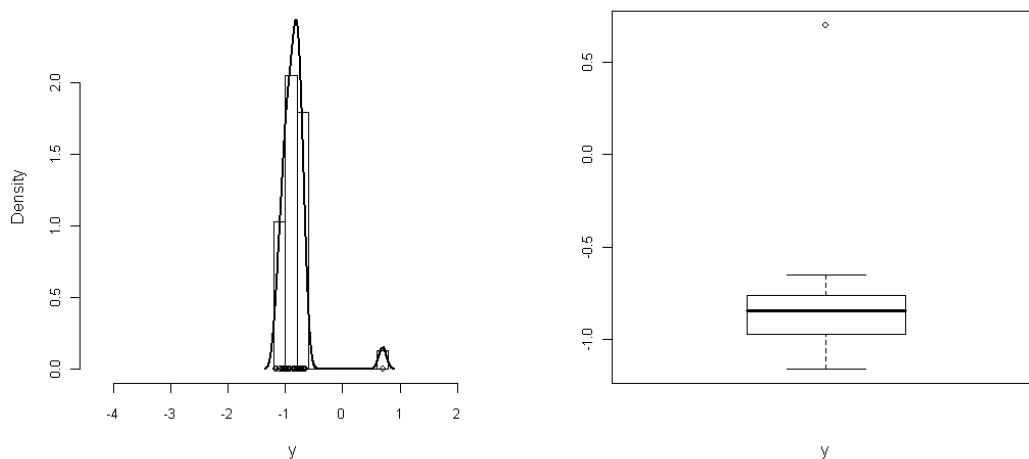


Figura 2.1 Pozzi 8-10 Istogramma con curva di densità non parametrica stimata (a sinistra), e diagramma con scatola e baffi (a destra).

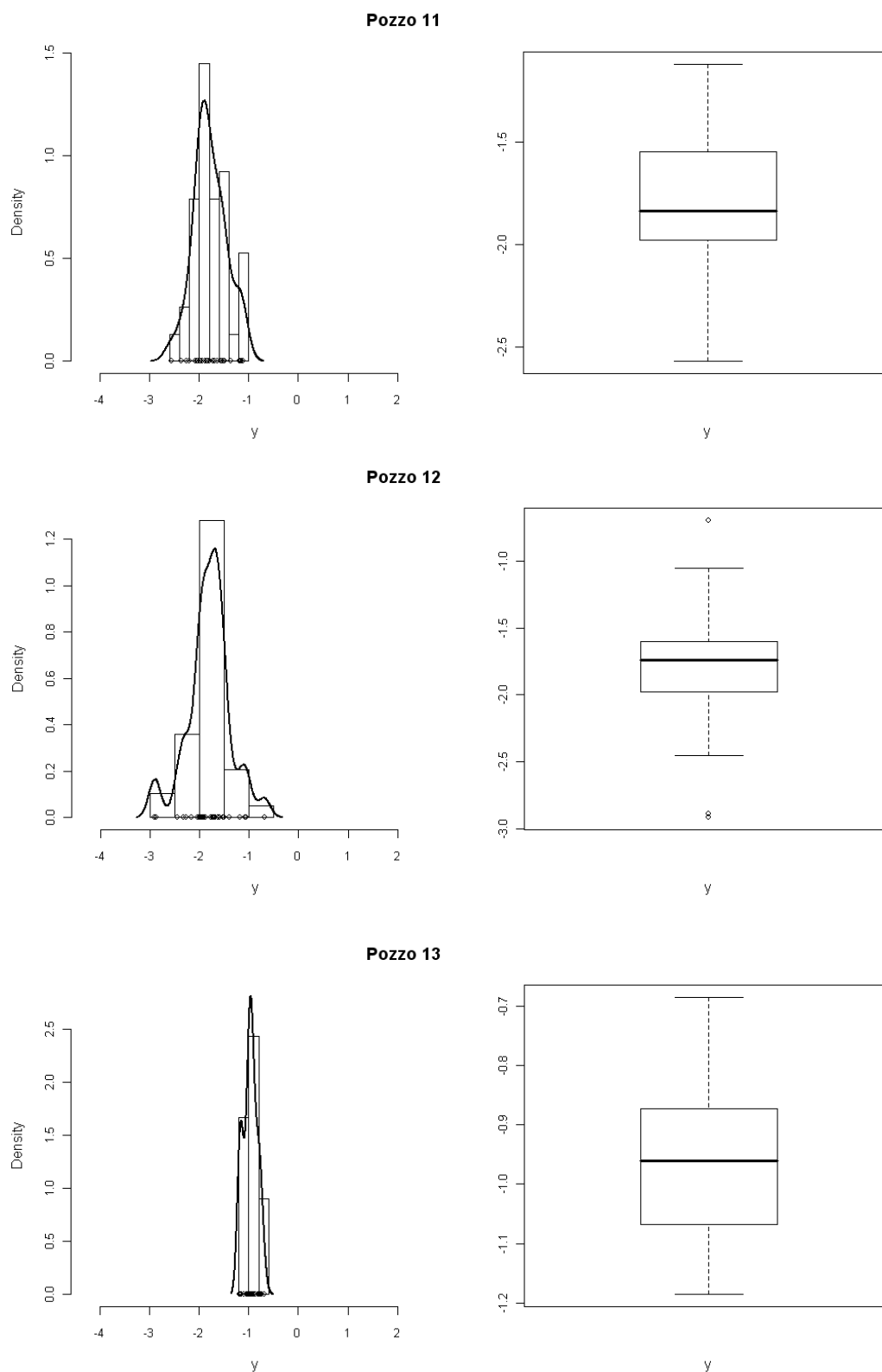
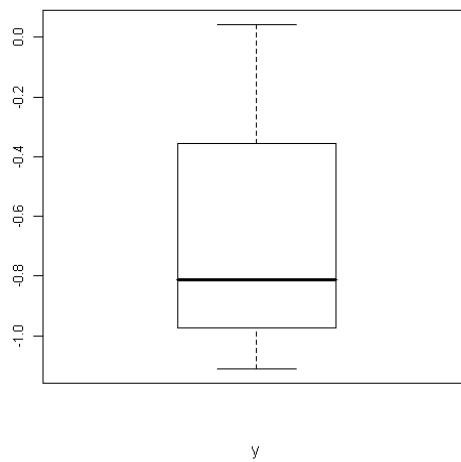
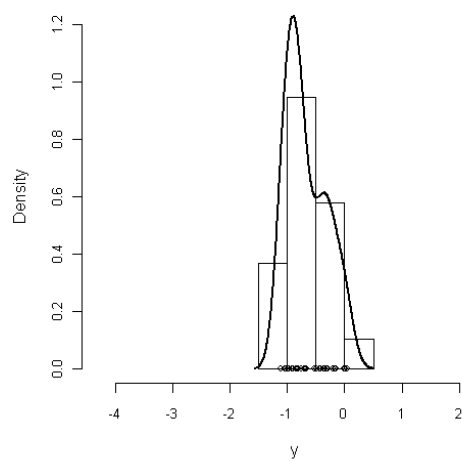
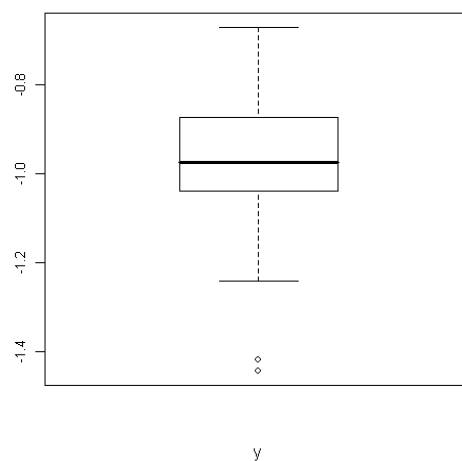
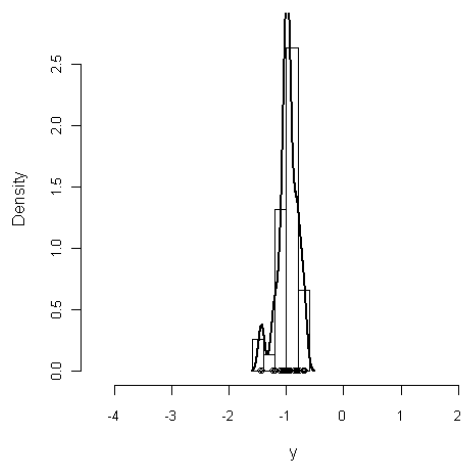


Figura 2.1 Pozzi 11-13 Istogramma con curva di densità non parametrica stimata (a sinistra), e diagramma con scatola e baffi (a destra).

Pozzo 14



Pozzo 15



Pozzo 16

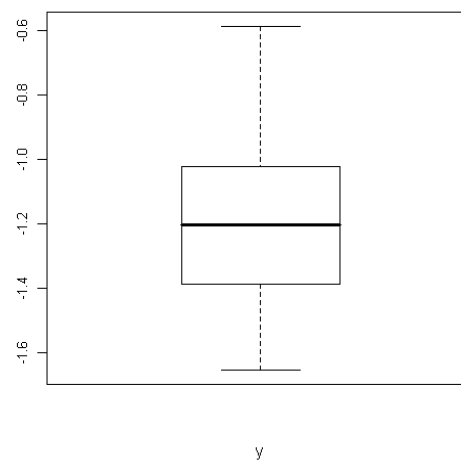
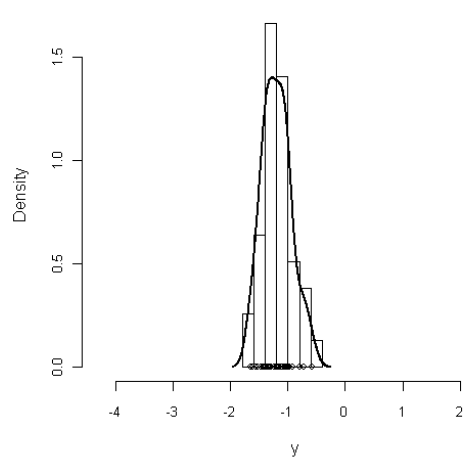
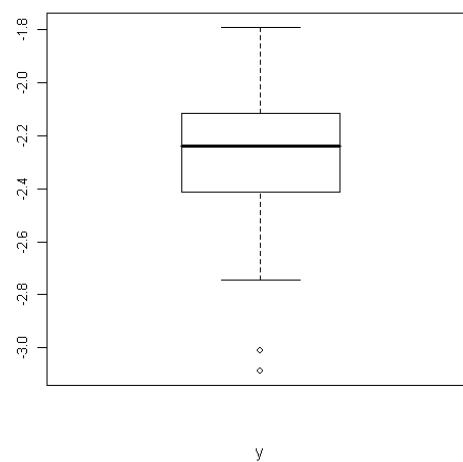
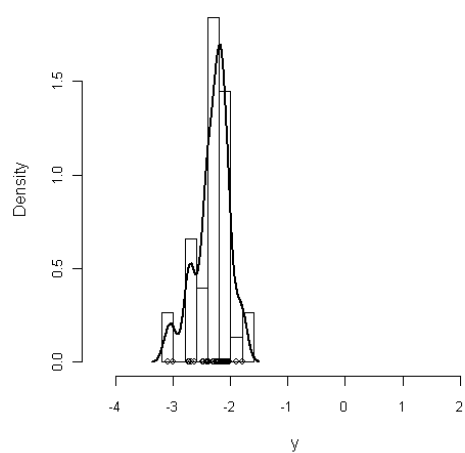
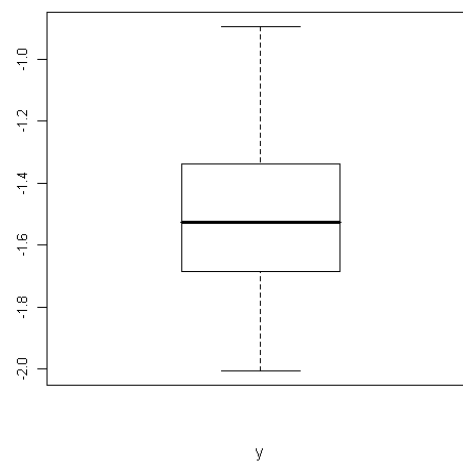
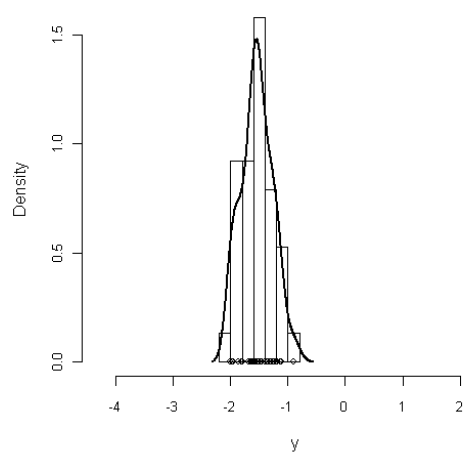


Figura 2.1 Pozzi 14-16 Istogramma con curva di densità non parametrica stimata (a sinistra), e diagramma con scatola e baffi (a destra).

Pozzo 17



Pozzo 18



Pozzo 19

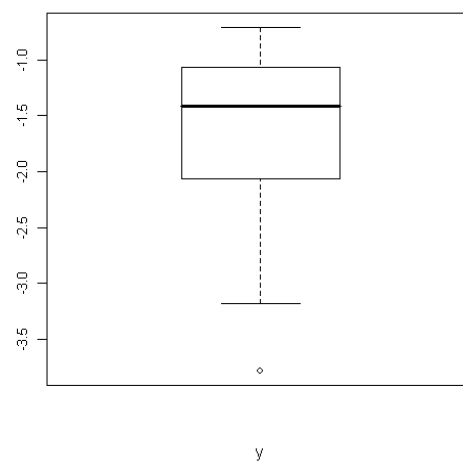
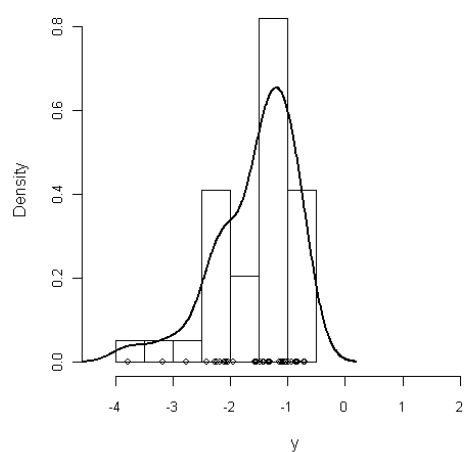
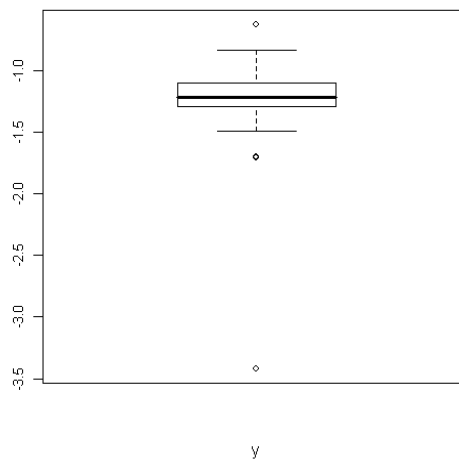
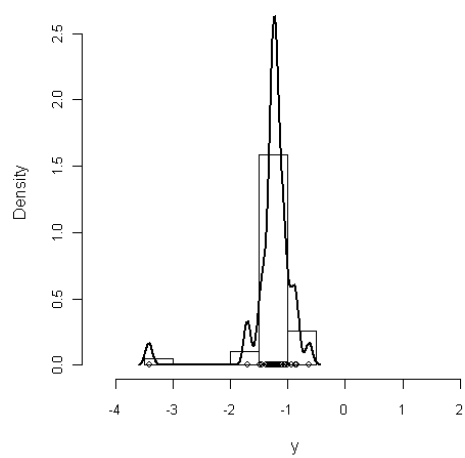
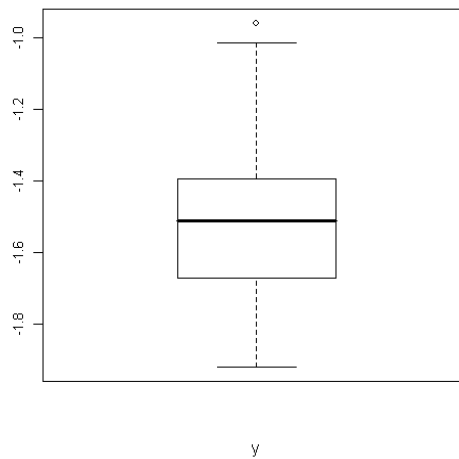
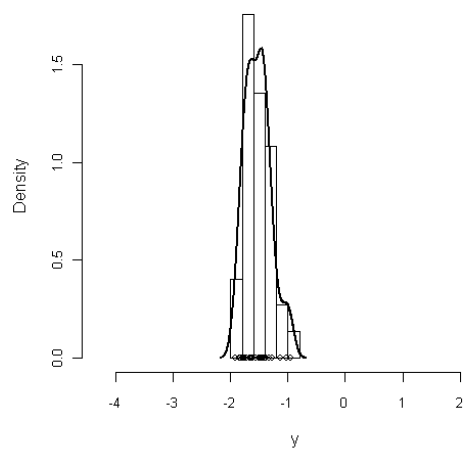


Figura 2.1 Pozzi 17-19 Istogramma con curva di densità non parametrica stimata (a sinistra), e diagramma con scatola e baffi (a destra).

Pozzo 20



Pozzo 21



Pozzo 22

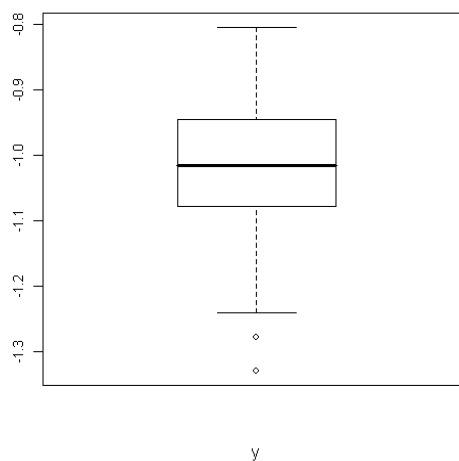
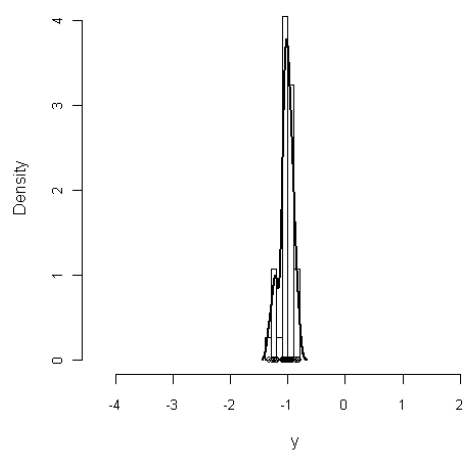


Figura 2.1 Pozzi 20-22 Istogramma con curva di densità non parametrica stimata (a sinistra), e diagramma con scatola e baffi (a destra).

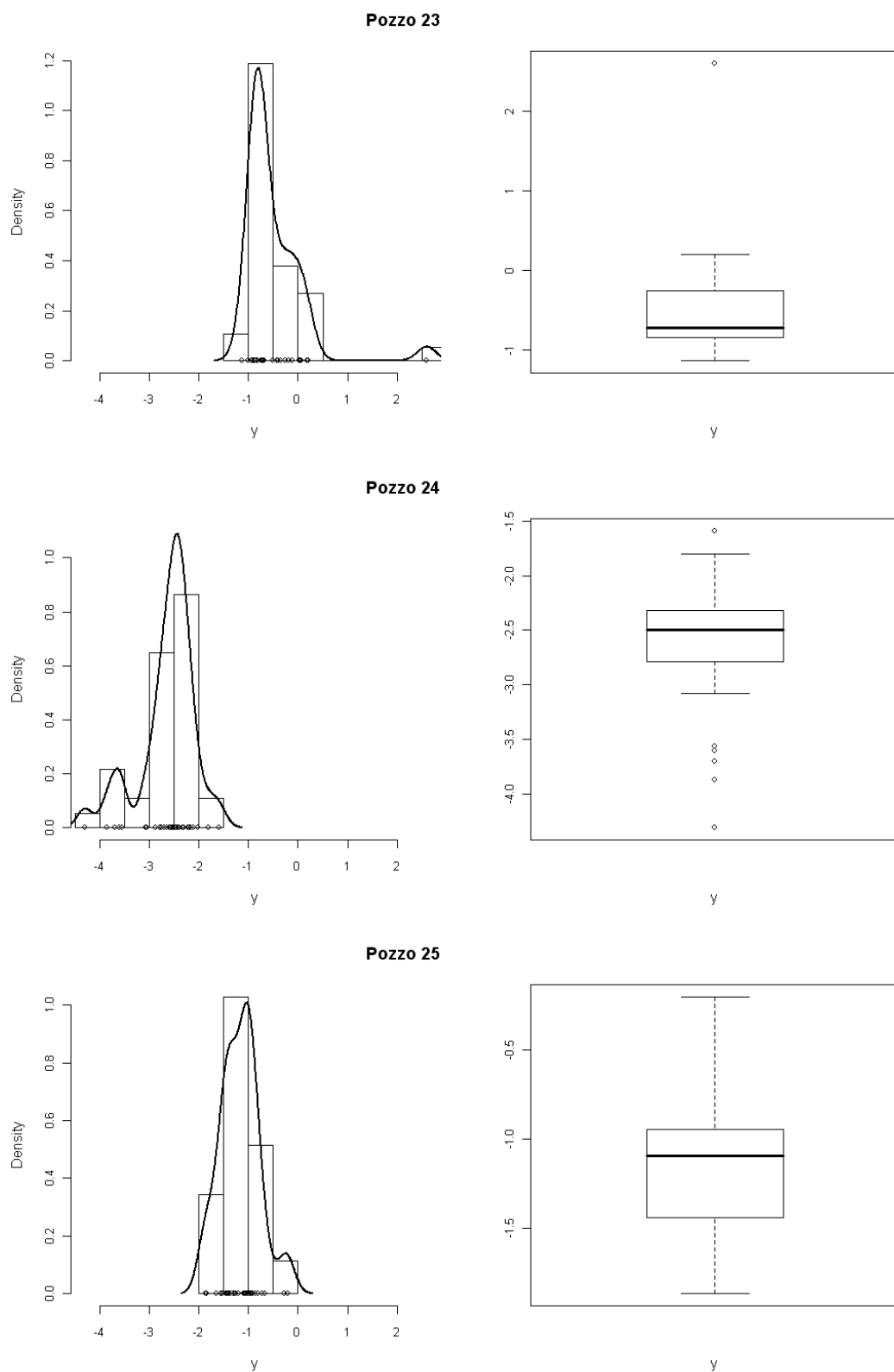


Figura 2.1 Pozzi 23-25 Istogramma con curva di densità non parametrica stimata (a sinistra), e diagramma con scatola e baffi (a destra).

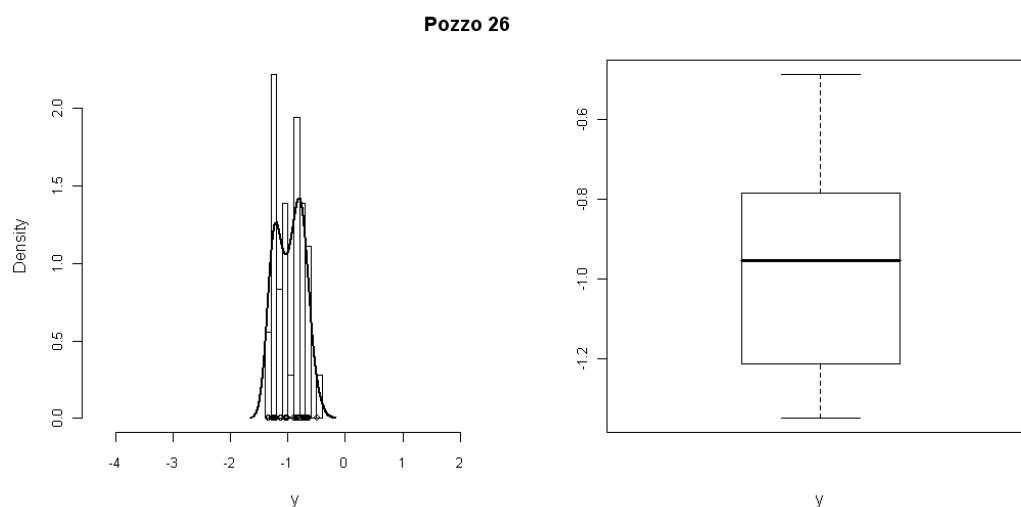


Figura 2.1 Pozzo 26 Istogramma con curva di densità non parametrica stimata (a sinistra), e diagramma con scatola e baffi (a destra).

Analizzando le rappresentazioni grafiche riportate, è possibile individuare alcune caratteristiche di tale variabile all'interno di ogni pozzo.

Per prima cosa, gli istogrammi sono stati creati imponendo la stessa ampiezza per l'asse delle ascisse; in questo modo risulta più semplice effettuare confronti tra i pozzi in base al campo di variabilità delle osservazioni. In particolare, è possibile individuare tre diversi comportamenti:

- le osservazioni sono contenute in un intervallo molto ristretto (ad esempio i pozzi 3 e 13);
- la maggior parte delle osservazioni è concentrata ma alcuni dati si discostano dagli altri in modo evidente (pozzi 2 e 10);
- il campo di variabilità è ampio e le osservazioni si distribuiscono in esso in tutta la sua larghezza (pozzi 7 e 9).

Si ha conferma di tali caratteristiche anche dai diagrammi con scatole e baffi (o *boxplot*) osservando l'ampiezza della scatola (determinata dalla distanza tra la prima e la terza linea) e la presenza di osservazioni estreme rappresentate da punti al di fuori dei baffi.

Come si è già accennato, la variabile assume valori sia positivi che negativi, ma dai grafici è possibile notare che il numero di osservazioni che superano lo zero sono poche, solo 18 su 977.

Gli strumenti utilizzati portano a risultati che permettono di identificare anche al-

cune caratteristiche della distribuzione da cui provengono le osservazioni. Infatti, gli istogrammi evidenziano un andamento a forma di campana che ricorda quello della distribuzione normale. Tuttavia emerge in maniera più o meno accentuata la presenza di asimmetria, positiva o negativa, che viene confermata anche dalla curva di densità non parametrica stimata sovrapposta all'istogramma. Tale andamento è riscontrabile anche dai diagrammi con scatole e baffi. Da questi è possibile valutare l'asimmetria, verificando la posizione non centrale della linea interna alla scatola in corrispondenza alla mediana, e la pesantezza delle code, indicata dall'estensione dei baffi.

Dalle considerazioni fatte è opportuno quindi condurre le analisi successive utilizzando una distribuzione con caratteristiche e andamento simili a quello della normale e che tenga conto dell'asimmetria.

Capitolo 3

LA DISTRIBUZIONE NORMALE ASIMMETRICA

Come accennato nel paragrafo precedente, i dati presentano andamento simile a quello della distribuzione normale, ma si discostano da questa per la presenza di asimmetria nelle osservazioni. La distribuzione normale asimmetrica è una distribuzione che soddisfa tali requisiti in quanto rappresenta un'estensione della densità normale con l'aggiunta di un parametro (parametro di forma) che ne regola appunto la forma.

3.1 Cenni preliminari

Una variabile casuale Z ha distribuzione normale asimmetrica se è continua con funzione di densità di probabilità

$$f(z) = 2\phi(z)\Phi(\alpha z), \quad (z \in \mathbb{R})$$

dove $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$ è la funzione di densità della Normale usuale,

$\Phi(\alpha z) = \int_{-\infty}^{\alpha z} \phi(t)dt$ è la sua funzione di distribuzione e α è il parametro che rego-

la l'asimmetria e può assumere valori nell'intervallo $(-\infty, \infty)$.

Quando il parametro di forma α risulta pari a zero l'asimmetria svanisce e si riottiene la densità $N(0,1)$; per $\alpha \rightarrow \infty$, invece, $f(z)$ converge alla seminormale.

Per introdurre i parametri di posizione e di dispersione, che finora non sono stati esaminati, si considera la trasformazione lineare

$$Y = \xi + \omega Z$$

che ha distribuzione normale asimmetrica con parametri (ξ, ω, α) e viene indicata con

$$Y \sim \text{SN}(\xi, \omega^2, \alpha)$$

dove ξ e ω sono rispettivamente, i parametri di posizione e di scala; ω si assume che sia positivo (Azzalini, A. e Capitanio, A., 1999).

Si noti che quando il parametro di asimmetria risulta $\alpha = 0$ si ottiene nuovamente la distribuzione $N(\xi, \omega^2)$.

3.2 Stima della distribuzione normale asimmetrica

Le analisi che seguono sono state svolte utilizzando il programma per la manipolazione dei dati *R*, in cui tutte le funzioni sono memorizzate in pacchetti. Quelle che fanno riferimento alla normale asimmetrica sono contenute nel pacchetto *sn*. In esso, sono previste due funzioni che, tramite la stima di massima verosimiglianza, permettono di ottenere una stima dei parametri (ξ, ω, α) della distribuzione. La prima, denominata *sn.mle*, è utilizzabile nei casi in cui la variabile osservata si presenta sotto forma di vettore; l'altra, denominata *msn.mle*, estende l'analisi al caso multidimensionale (la variabile osservata può essere vettore o matrice). L'oggetto di studio è un vettore che descrive il logaritmo naturale del rapporto tra catione calcio e anione solfato (indicato nei grafici tramite la variabile *y*) quindi, per stimare i parametri della distribuzione di ogni pozzo, possono essere

utilizzate entrambe le funzioni, che nonostante eseguano operazioni diverse, ci si aspetta diano le medesime stime. Seguendo tale procedura, però, sono state registrate alcune differenze tra i due risultati. Per verificare tale situazione, per ogni pozzo vengono riportate le curve stimate tramite le due funzioni (la linea tratteggiata indica la stima ottenuta con la funzione *sn.mle*, la linea continua indica la stima ottenuta con la funzione *msn.mle*) e i valori dei parametri della distribuzione.

Graficamente è possibile notare se le due funzioni hanno portato a risultati diversi verificando se le rispettive curve coincidono o si differenziano l'una dall'altra.

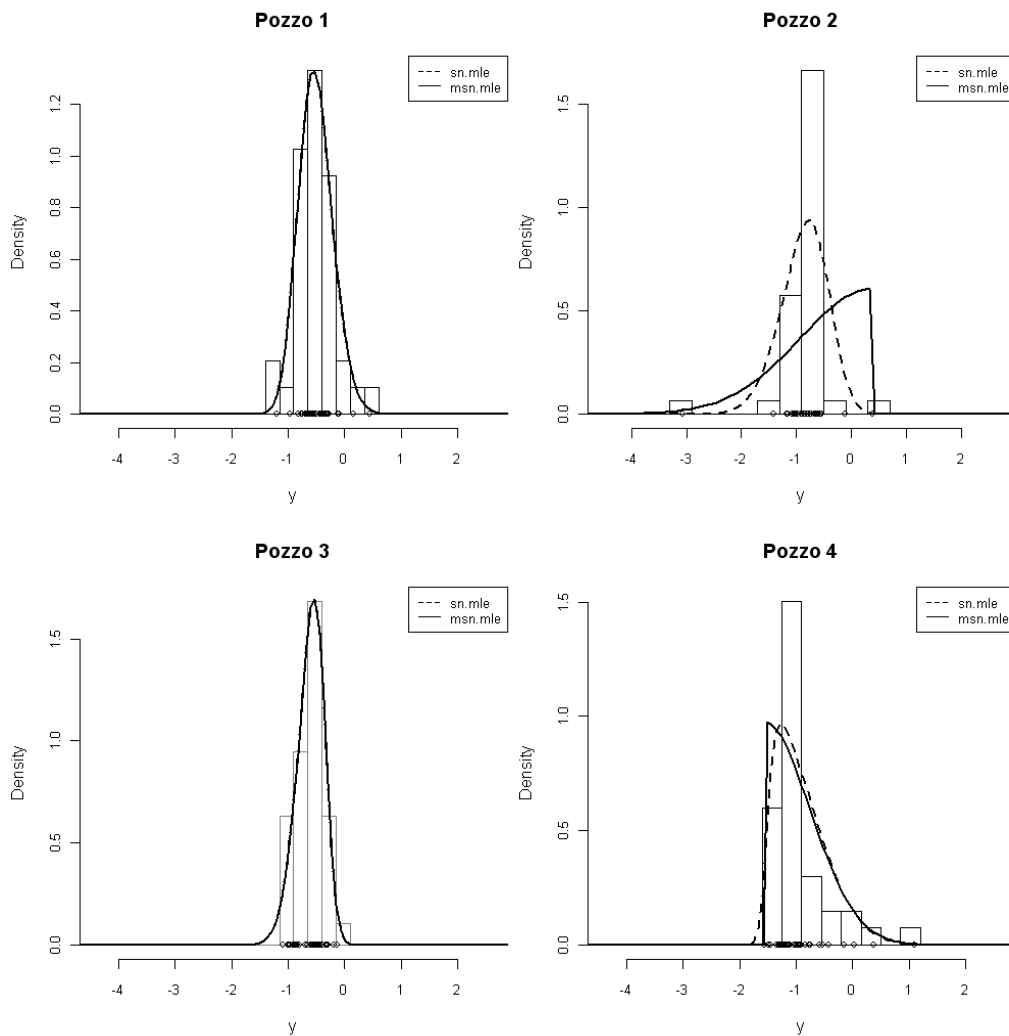


Figura 3.1 Pozzi 1-4 Istogramma con le curve della densità normale asimmetrica stimate tramite *sn.mle* (linea tratteggiata) e tramite *msn.mle* (linea continua).

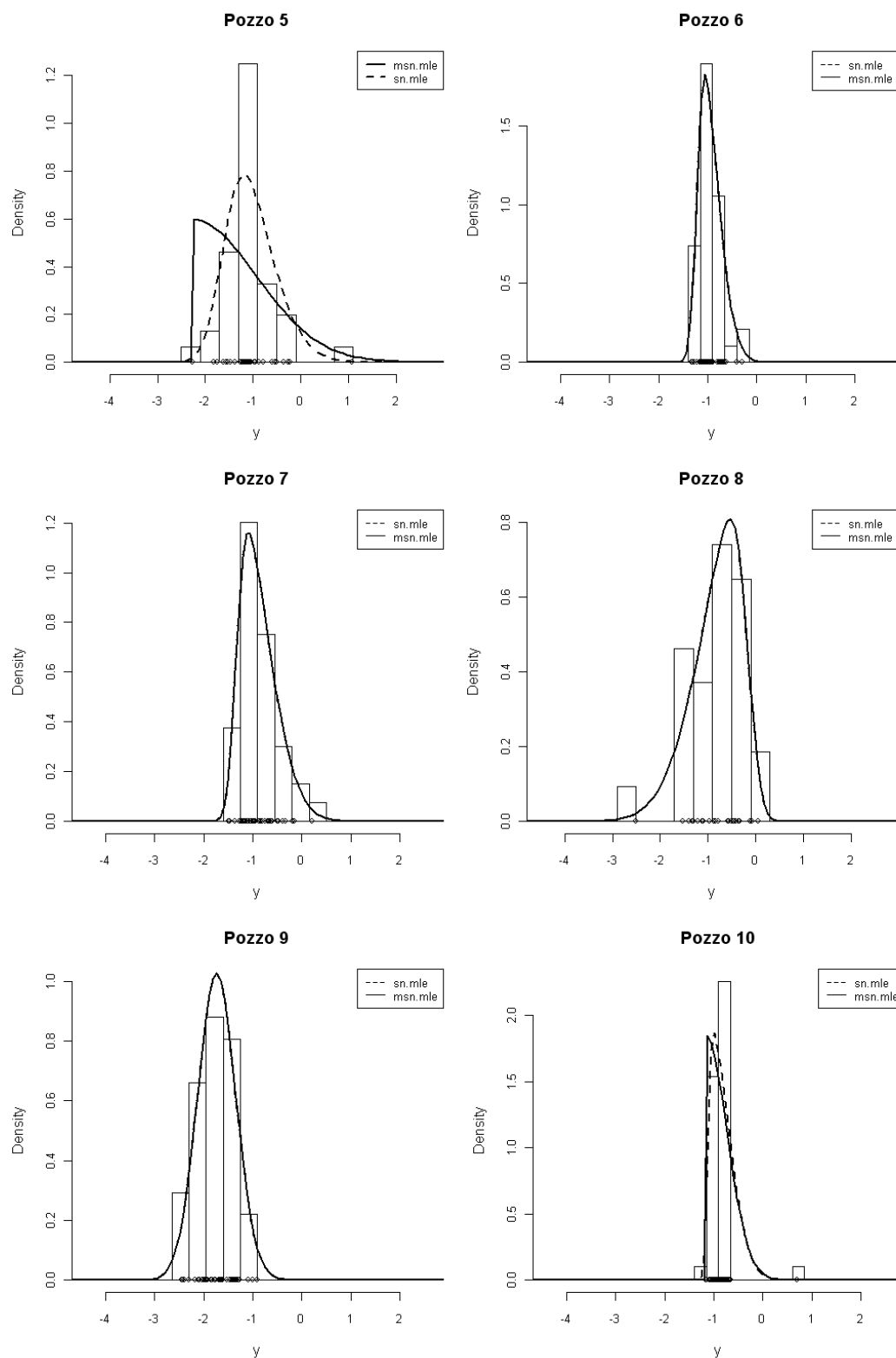


Figura 3.1 Pozzi 5-10 Istogramma con le curve della densità normale asimmetrica stimate tramite *sn.mle* (linea tratteggiata) e tramite *msn.mle* (linea continua).

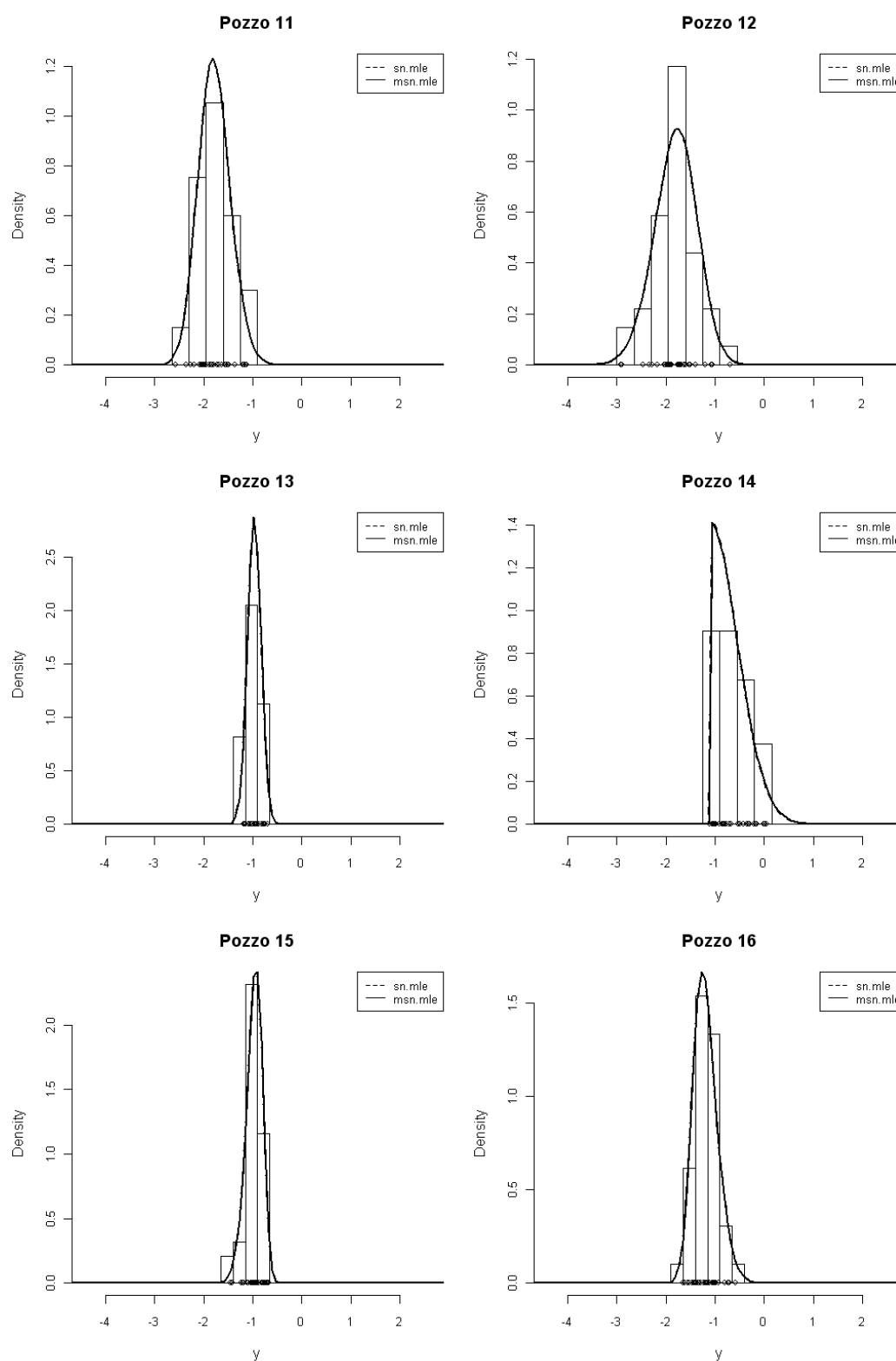


Figura 3.1 Pozzi 11-16 Istogramma con le curve della densità normale asimmetrica stimate tramite *sn.mle* (linea trattteggiata) e tramite *msn.mle* (linea continua).

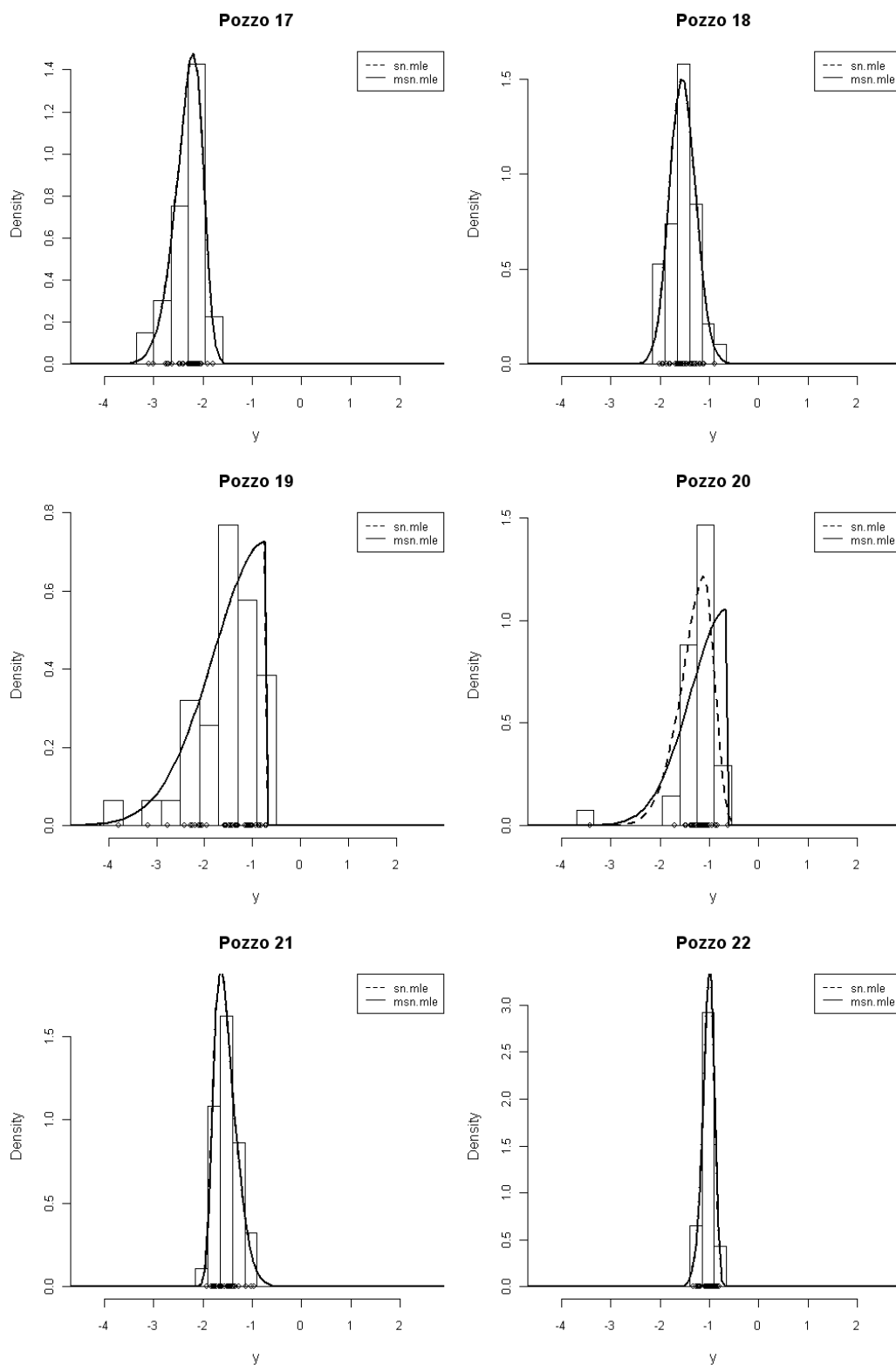


Figura 3.1 Pozzi 17-22 Istogramma con le curve della densità normale asimmetrica stimate tramite *sn.mle* (linea tratteggiata) e tramite *msn.mle* (linea continua).

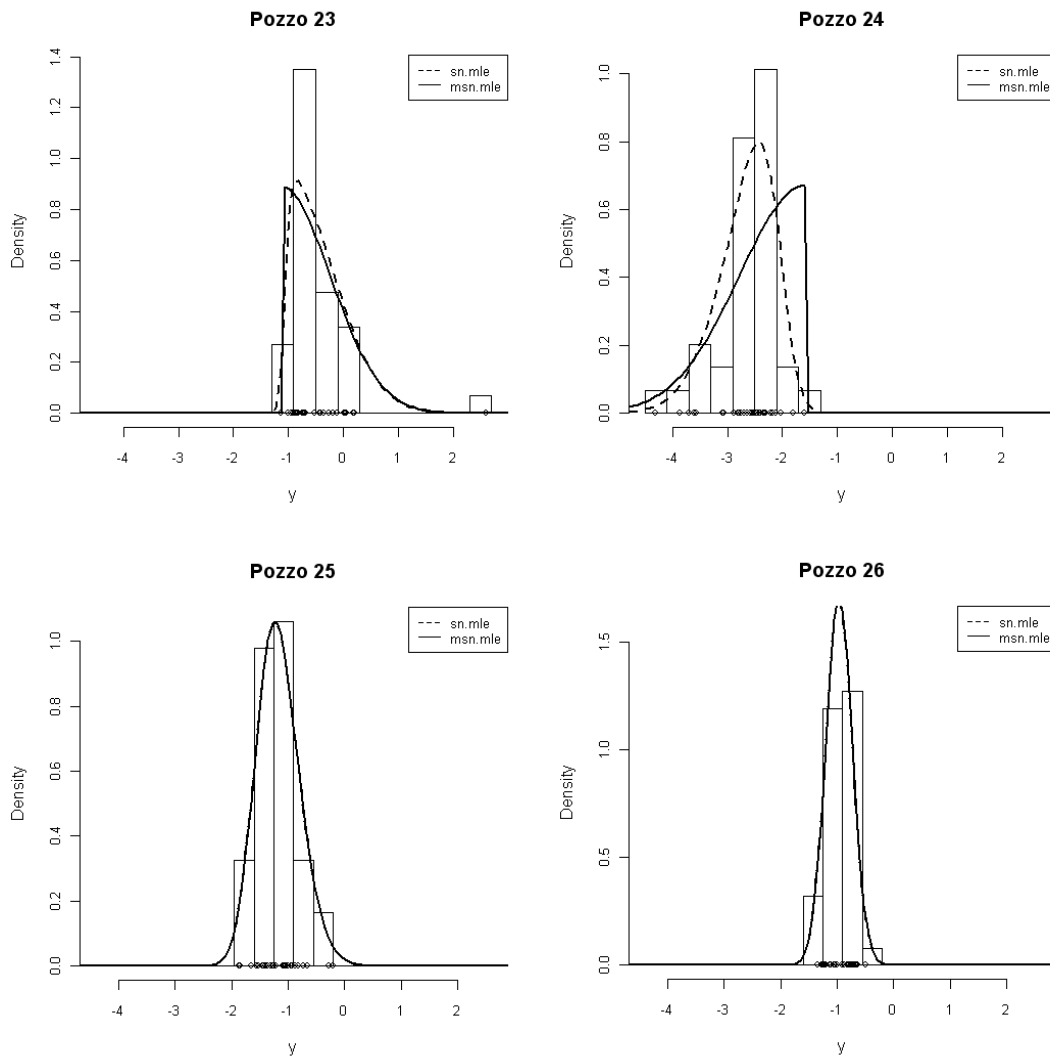


Figura 3.1 Pozzi 23-26 Istogramma con le curve della densità normale asimmetrica stimate tramite *sn.mle* (linea tratteggiata) e tramite *msn.mle* (linea continua).

Il primo aspetto che risulta evidente dai grafici è l'accordo tra istogramma e curva di densità parametrica stimata: risulta nettamente migliore rispetto a quello ottenuto utilizzando la densità non parametrica nel paragrafo 2.2.

Si nota però che per alcuni pozzo le curve stimate secondo i due metodi non coincidono ma presentano alcune diversità. Per verificare tale aspetto di seguito vengono riportati i valori dei parametri stimati supponendo che ogni pozzo Y abbia distribuzione normale asimmetrica con parametri (ξ, ω, α) , dove ξ , ω e α indicano rispettivamente i parametri di posizione, di scala e di forma.

Pozzo	ξ	ω	α	logL
1	-0.763	0.396	1.358	-8.672
2	0.388	1.312	-1.58x10 ⁷	-38.898
3	-0.357	0.352	-2.076	0.140
4	-1.566	0.816	8.35x10 ⁶	-19.861
5	-2.266	1.334	1.45x10 ⁷	-38.549
6	-1.219	0.365	3.201	2.541
7	-1.337	0.599	4.067	-14.644
8	-0.169	0.855	-3.851	-20.229
9	-1.879	0.415	0.490	-18.558
10	-1.161	0.431	3.06x10 ⁷	4.508
11	-2.033	0.413	1.182	-11.427
12	-1.474	0.549	-1.185	-22.808
13	-0.983	0.139	0.138	21.750
14	-1.112	0.562	9.35x10 ⁵	-5.739
15	-0.803	0.241	-2.025	14.551
16	-1.448	0.355	2.011	-0.389
17	-1.989	0.432	-2.681	-5.244
18	-1.709	0.316	0.923	-3.570
19	-0.706	1.095	-2.48x10 ⁶	-31.877
20	-0.628	0.754	-1.74x10 ⁷	-17.309
21	-1.798	0.354	3.562	4.274
22	-0.896	0.176	-2.526	27.615
23	-1.131	0.894	6.97x10 ⁵	-22.712
24	-1.589	1.190	-1.00x10 ⁸	-33.296
25	-1.509	0.511	1.481	-15.881
26	-0.981	0.233	0.083	1.383

(b) Stime dei parametri ottenute con la funzione *msn.mle* e valore della log-verosimiglianza nel massimo.

Pozzo	ξ	ω	α	logL
1	-0.763	0.396	1.358	-8.672
2	-0.439	0.611	-1.809	-22.226
3	-0.357	0.352	-2.076	0.140
4	-1.509	0.773	7.079	-21.278
5	-1.584	0.765	2.111	-28.900
6	-1.219	0.365	3.201	2.541
7	-1.337	0.599	4.067	-14.644
8	-0.169	0.855	-3.851	-20.229
9	-1.878	0.415	0.490	-18.558
10	-1.118	0.399	6.945	3.943
11	-2.033	0.413	1.182	-11.427
12	-1.474	0.549	-1.185	-22.808
13	-0.983	0.139	0.138	21.750
14	-1.113	0.556	1.53x10 ³	-5.804
15	-0.803	0.241	-2.025	14.551
16	-1.448	0.355	2.011	-0.389
17	-1.989	0.432	-2.681	-5.244
18	-1.709	0.316	0.923	-3.570
19	-0.703	1.096	-1.53x10 ³	-31.972
20	-0.884	0.557	-3.475	-12.497
21	-1.798	0.354	3.562	4.274
22	-0.896	0.176	-2.526	27.615
23	-1.054	0.840	9.837	-22.727
24	-2.039	0.820	-2.981	-27.868
25	-1.509	0.511	1.481	-15.881
26	-0.981	0.233	0.083	1.383

(a) Stime dei parametri ottenute con la funzione *sn.mle* e valore della log-verosimiglianza nel massimo.

Tabella 3.1 Riassunto dei valori ottenuti tramite i due metodi di stima considerati.

Sia dalle rappresentazioni grafiche che dai valori riportati nelle tabelle, si nota che per nove pozzi (2, 4, 5, 10, 14, 19, 20, 23, 24) le due funzioni utilizzate hanno portato a stime differenti. In particolare, in tali casi, le due curve disegnate non coincidono, il valore della log-verosimiglianza nei punti di massimo indicati è diversa e, tra le stime dei parametri, la discrepanza maggiore è nell'indice di asimmetria che, con l'utilizzo della funzione *msn.mle*, assume va-

lori che tendono a infinito (e convergono quindi alla distribuzione seminormale). Quest'ultimi rappresentano casi in cui la stima è qualitativamente poco soddisfacente. Per ovviare a questo problema, dato che la funzione *sn.mle* presenta un numero superiore di stime dell'indice di forma con valore finito, la log-verosimiglianza assume valore superiore nei punti di massimo stimati da questa funzione (ad eccezione dei pozzi 4, 10, 14, 19, 23) e graficamente le curve delle distribuzioni risultano seguire meglio i dati, al fine di ottenere risultati più affidabili, nel seguito si farà riferimento a tale metodo di stima (questo aspetto verrà analizzato in maniera più approfondita nel paragrafo 3.3).

Nonostante la funzione *sn.mle* sia stata considerata più adeguata, la sua applicazione nei pozzi 14 e 19 ha portato a stime del parametro di forma con valori che tendono a ∞ . Per poter ottenere valori finiti viene utilizzato un metodo di stima, contenuto nella funzione *sn.mmle* e applicabile quando la variabile oggetto di studio è scalare, secondo il quale la procedura di massimizzazione viene riavviata e fermata quando raggiunge un valore della log-verosimiglianza che non è in maniera significativa inferiore al massimo. In questo modo i parametri e i grafici delle curve di densità così stimati risultano:

Pozzo	ξ	ω	α
14	-1.1130	0.556	8.762
19	-0.703	1.096	-7.606

Tabella 3.2 Stime dei parametri ottenute con la funzione *sn.mmle*.

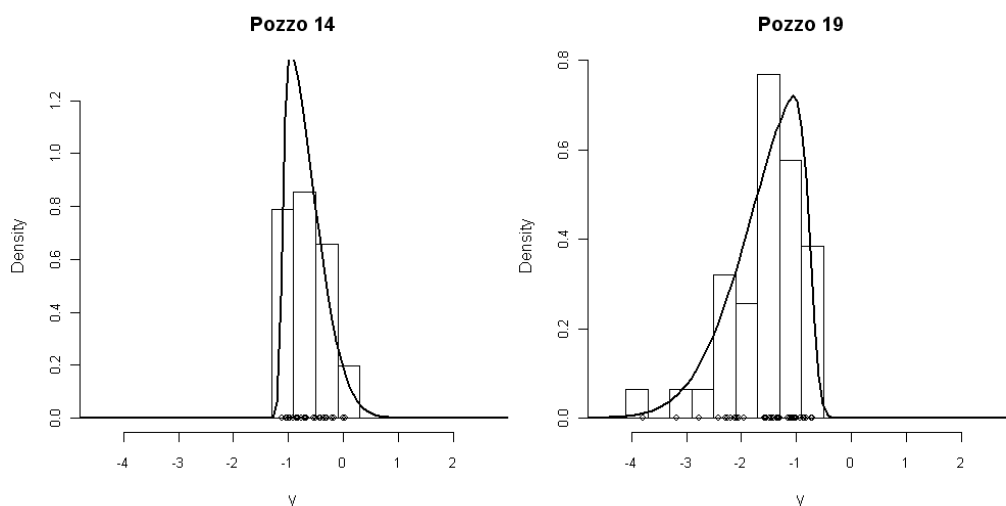


Figura 3.2 Istogramma con le curve della densità normale asimmetrica stimate tramite *sn.mmle* per i pozzi 14, 19.

Nel complesso, dalle analisi svolte si ha conferma della presenza di asimmetria sia negativa che positiva. Il primo caso si riscontra in dieci pozzi la cui stima di α assume valore inferiore a zero, di conseguenza la coda sinistra assume peso maggiore e l'asimmetria è negativa (la più accentuata nel pozzo 8). Nei rimanenti sedici pozzi ve ne sono quattro che presentano forma a campana simmetrica: il loro parametro di asimmetria è positivo ma vicino a zero (pozzi 9, 13, 18, 26).

In base alle considerazioni fatte nel paragrafo 2.2, dalle quali è emerso che i dati richiedono una distribuzione con andamento simile a quello della normale ma con asimmetria, e dalle analisi appena svolte, in cui è possibile riscontrare dalle rappresentazioni grafiche un accordo soddisfacente tra andamento dei dati e distribuzione stimata, si può confermare che la distribuzione normale asimmetrica è la distribuzione parametrica adeguata per spiegare il logaritmo naturale del rapporto tra calcio e solfato.

3.3 Aspetti salienti della log-verosimiglianza

Nel paragrafo precedente, a seconda della funzione scelta per la stima dei parametri della distribuzione, per alcuni pozzi si sono verificate incongruenze tra i risultati. In un primo momento è stato scelto di considerare la funzione *sn.mle* più affidabile in quanto presenta un numero superiore di stime dell'indice di forma con valore finito e graficamente le rappresentazioni delle relative distribuzioni risultano seguire meglio i dati. In questo paragrafo invece, si vuole verificare se tale funzione permette di ottenere risultati più attendibili ed è adeguata per le analisi successive.

Innanzitutto, si deve tenere conto che le funzioni utilizzate si basano sul metodo della stima di massima verosimiglianza: il risultato ottenuto rappresenta un punto di massimo della log-verosimiglianza. Nei casi in cui tale funzione non presenti forma quadratica regolare ma vi siano più punti di massimo locale, per verificare se quello indicato dalla funzione è effettivamente il massimo assoluto si deve accertare che il valore della log-verosimiglianza assuma valore maggiore rispetto a-

gli altri massimi locali.

Da queste considerazioni è possibile ipotizzare che le differenze riscontrate nelle stime dei parametri siano dovute alla presenza di vari massimi locali nella funzione di log-verosimiglianza profilo di α . Le analisi che seguono hanno l'obiettivo, quindi, di verificarne l'esistenza e confermare che il punto indicato dalla funzione *sn.mle* sia il massimo assoluto e assuma valore della log-verosimiglianza più alto. Vengono presi in esame i pozzi 2, 4, 5, 10, 14, 19, 20, 23, 24. Per ognuno di essi si considerano le stime ottenute con entrambe le funzioni analizzando la log-verosimiglianza profilo di α e del parametro bidimensionale (ω, α) ².

In particolare vengono riportati quattro grafici³:

- (a) - (b) analizzano rispettivamente la log-verosimiglianza profilo di α e di (ω, α) verificando i risultati ottenuti dalla funzione *sn.mle*;
- (c) - (d) analizzano la log-verosimiglianza profilo di α prima in un intervallo ampio che comprende entrambi i valori indicati dalle due funzioni e poi in un intervallo che include solo il massimo indicato dalla funzione *msn.mle*.

• Pozzo 2

La log-verosimiglianza profilo relativa in Figura 3.3 (a) presenta una forma molto vicina a quella parabolica e, come accennato, assume valore zero nel suo punto di massimo (indicato con il simbolo “*”) e il cui corrispondente valore di α coincide con quello stimato tramite la funzione *sn.mle*. Anche nelle curve di livello in Figura 3.3 (b) la regione di confidenza, con quantile di livello 0.25 della distribuzione χ^2_2 , contiene il punto di coordinate i valori stimati ($\hat{\omega} = 0.611, \hat{\alpha} = -1.809$).

Il grafico (c) mostra il comportamento della log-verosimiglianza profilo relativa in un intervallo dell'asse delle ascisse che include entrambi i massimi indicati dalle due funzioni ($-18'000'000 \leq \alpha \leq 0$); in questo è possibile verificare che per

² Più precisamente, si considera la log-verosimiglianza profilo relativa che assume solo valori non positivi; se ha forma simile a quella quadratica nel punto di massimo, che coincide con la stima di massima verosimiglianza, la funzione assume valore zero.

³ Le rappresentazioni grafiche che seguono sono state create tramite la funzione *sn.2logL.profile*, anch'essa contenuta nel pacchetto *sn* di R.

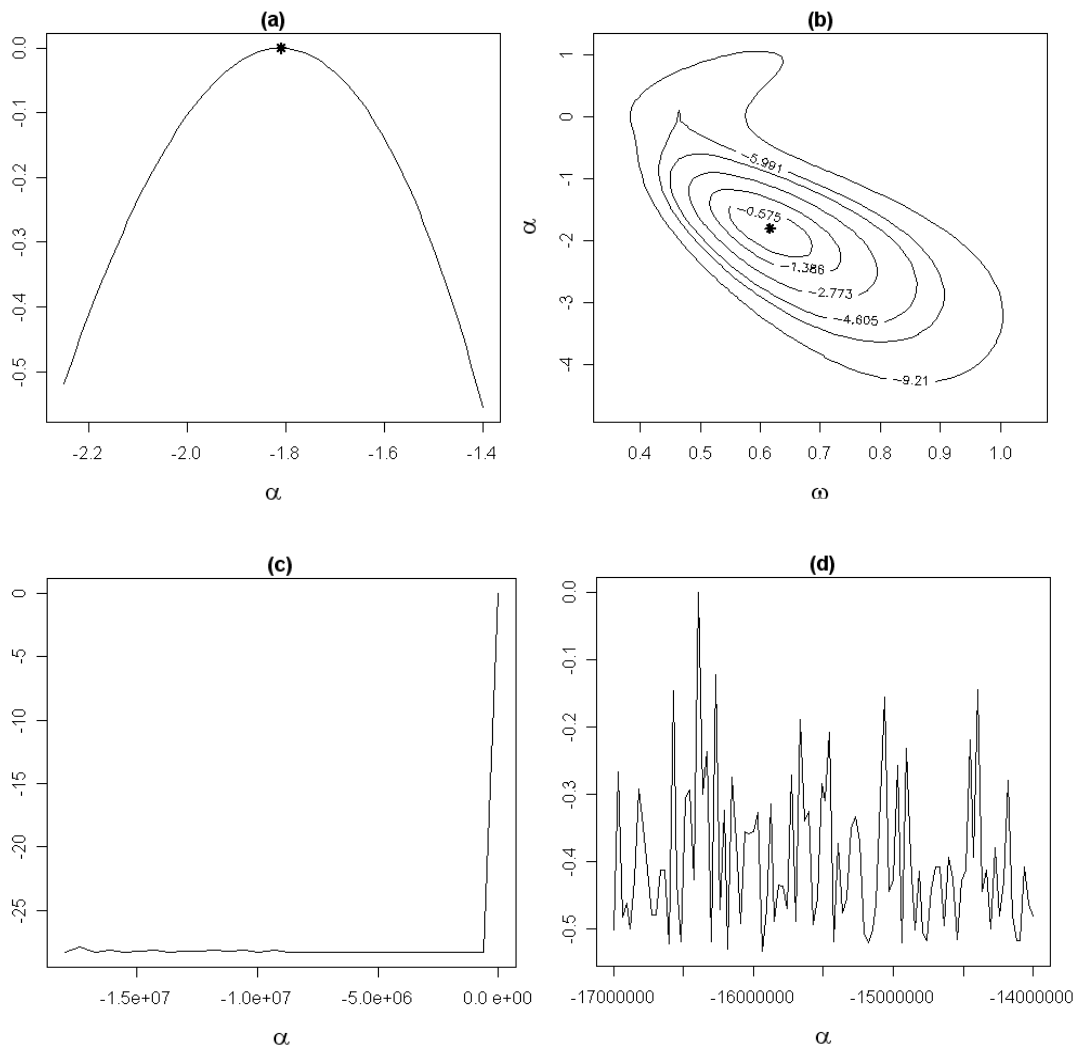


Figura 3.3 Pozzo 2: due volte la verosimiglianza profilo relativa di α e curve di livello della stessa funzione per (ω, α) verificando le stime ottenute tramite *sn.mle* (a-b) e due volte la log-verosimiglianza profilo relativa di α in intervalli più ampi (c-d).

$-2.2 \leq \alpha \leq -1.4$			$-17000000 \leq \alpha \leq -14000000$	
ω	α	$\log L$	α	$\log L$
0.611	-1.809	-22.226	-1.58×10^7	-39.219

Tabella 3.3 Pozzo 2: valore dei parametri e della log-verosimiglianza secondo i due intervalli considerati nei grafici (a) - (d).

$\alpha \rightarrow -\infty$ la funzione non è stazionaria ma presenta oscillazioni: ciò dimostra la presenza di più massimi locali. Infatti nel grafico (d), creato imponendo un intervallo che includesse il valore indicato dalla stima con *msn.mle*, è possibile notare che in un punto la funzione assume valore zero.

Ciò che ha maggiore rilevanza è che in corrispondenza di questo punto di massi-

mo locale la log-verosimiglianza assume valore pari a -39.219 ed, essendo inferiore a quello in corrispondenza di $\alpha = -1.809$ in cui $\log L = -22.226$, si ha conferma che la stima ottenuta con *sn.mle* è il punto di massimo assoluto.

• Pozzo 4

Lo stesso procedimento è stato seguito anche per l'analisi del pozzo 4, ma facendo particolare attenzione in quanto la log-verosimiglianza nel punto di massimo ottenuto con *msn.mle* è risultata superiore.

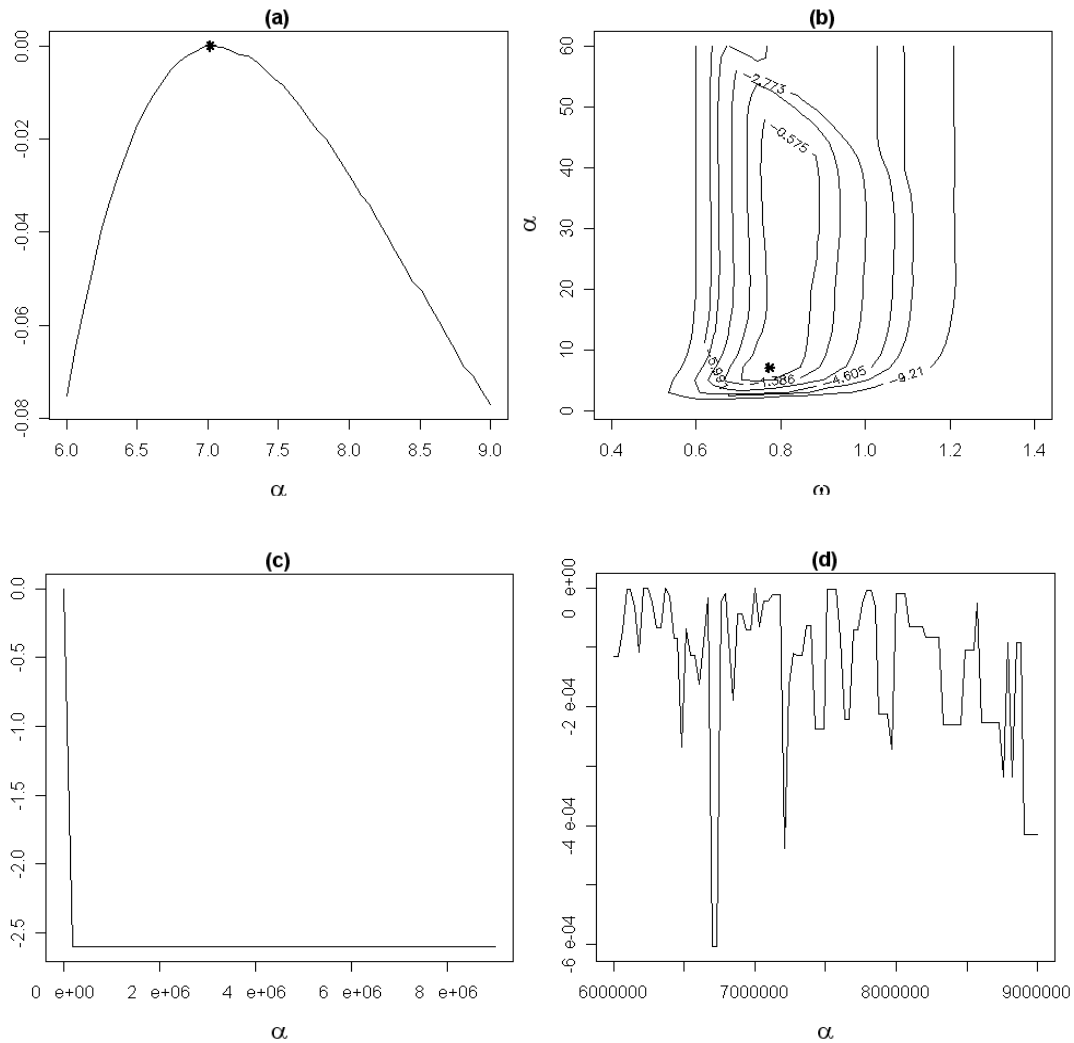


Figura 3.4 Pozzo 4: due volte la verosimiglianza profilo relativa di α e curve di livello della stessa funzione per (ω, α) verificando le stime ottenute tramite *sn.mle* (a-b) e due volte la log-verosimiglianza profilo relativa di α in intervalli più ampi (c-d).

$6 \leq \alpha \leq 9$			$6000000 \leq \alpha \leq 9000000$	
$\hat{\omega}$	$\hat{\alpha}$	$\log L$	α	$\log L$
0.773	7.079	-21.278	7.35×10^6	-22.786

Tabella 3.4 Pozzo 4: valore dei parametri e della log-verosimiglianza secondo i due intervalli considerati nei grafici (a) - (d).

Anche in questo caso i grafici della prima riga sono congruenti con le stime ottenute con la funzione *sn.mle*: la log-verosimiglianza profilo relativa nel punto stimato (indicato con il simbolo “*”) ha valore pari a zero e nelle curve di livello il punto ($\hat{\omega}=0.773, \hat{\alpha}=7.798$) si trova nella regione di confidenza di livello 0.25.

A differenza del pozzo 2, analizzando la Figura 3.4 (c), in cui si considera α nell’intervallo $5 \leq \alpha \leq 9\,000\,000$, non sono evidenti oscillazioni ma, esaminando il grafico (d), in cui la log-verosimiglianza profilo relativa è considerata nell’intervallo che comprende solo la stima ottenuta tramite *msn.mle*, $6\,000\,000 \leq \alpha \leq 9\,000\,000$, vi è la presenza di più punti di massimo locale.

In contrasto ai valori riportati in Tabella 3.1, tramite queste analisi viene confermato che, nonostante la differenza sia minima, la log-verosimiglianza assume valore superiore nel punto di massimo indicato dalla funzione *sn.mle* ($\log L=-21.28$) e più basso negli altri massimi locali (ad esempio, in prossimità di $\alpha=8.3 \times 10^6$, $\log L=-22.786$).

• Pozzo 5

Nei grafici (a) e (b) in Figura 3.5, che analizzano il pozzo 5, la log-verosimiglianza profilo relativa assume valore pari a zero nel punto indicato dalla funzione *sn.mle* ($\hat{\alpha}=2.111$) e nelle curve di livello il punto individuato dalle coordinate ($\hat{\omega}=0.765, \hat{\alpha}=2.111$) si trova nella regione con livello di confidenza 0.25, a conferma dei risultati ottenuti applicando questo metodo di stima.

Nel grafico (c) si considerano valori di α nell’intervallo $1 \leq \alpha \leq 15\,000\,000$ ed è possibile notare la presenza di oscillazioni confermate anche nel grafico (d); in quest’ultimo viene evidenziato un altro punto di massimo ma, in corrispondenza a questo, la log-verosimiglianza risulta pari a -38.549 e dimostra il massimo assoluto in $\alpha=2.111$ dove $\log L=-28.9$.

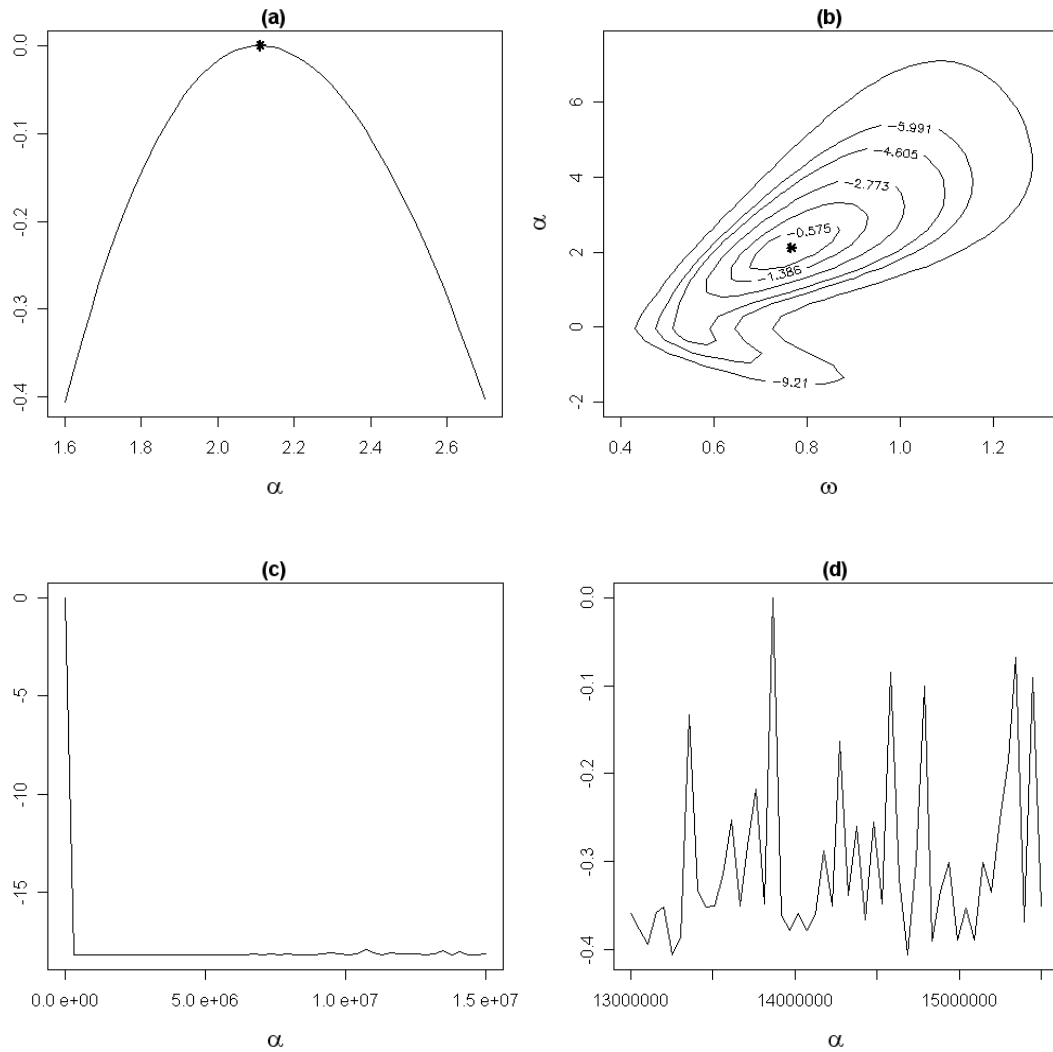


Figura 3.5 Pozzo 5: due volte la verosimiglianza profilo relativa di α e curve di livello della stessa funzione per (ω, α) verificando le stime ottenute tramite *sn.mle* (a-b) e due volte la log-verosimiglianza profilo relativa di α in intervalli più ampi(c-d).

$1.6 \leq \alpha \leq 2.6$			$13000000 \leq \alpha \leq 15000000$	
ω	α	$\log L$	α	$\log L$
0.765	2.111	-28.900	1.38×10^7	-38.549

Tabella 3.5 Pozzo 5: valore dei parametri e della log-verosimiglianza secondo i due intervalli considerati nei grafici (a) - (d).

• **Pozzo 10**

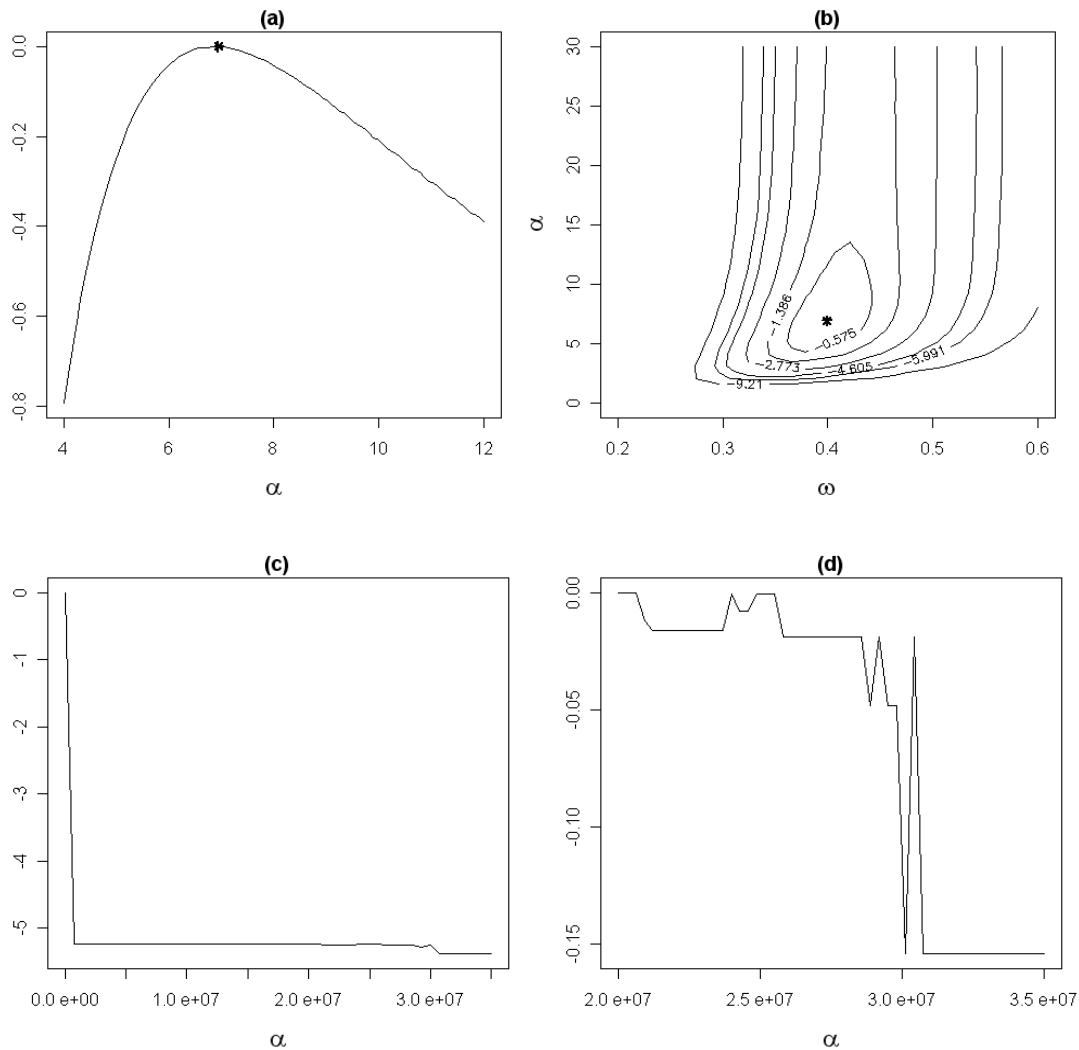


Figura 3.6 Pozzo 10: due volte la verosimiglianza profilo relativa di α e curve di livello della stessa funzione per (ω, α) verificando le stime ottenute tramite *sn.mle* (a-b) e due volte la log-verosimiglianza profilo relativa di α in intervalli più ampi (c-d).

$4 \leq \alpha \leq 2$			$20000000 \leq \alpha \leq 35000000$	
ω	α	$\log L$	α	$\log L$
0.399	6.945	3.943	2.06×10^7	1.301

Tabella 3.6 Pozzo 10: valore dei parametri e della log-verosimiglianza secondo i due intervalli considerati nei grafici (a) - (d).

Si riscontra ancora una corrispondenza tra i risultati ottenuti tramite la funzione *sn.mle* e i grafici (a) e (b) riportati in Figura 3.6.

Nel riquadro (c), invece, vengono considerati valori di α nell'intervallo $6 \leq \alpha \leq 35000000$ ed è possibile notare alcune oscillazione per $\alpha \rightarrow +\infty$; il gra-

fico (d) (in cui $20'000'000 \leq \alpha \leq 35'000'000$) conferma la presenza di un punto di massimo locale in corrispondenza del quale, a differenza del valore riportato in Tabella 3.1, la log-verosimiglianza ora risulta pari a 1.301, valore inferiore a quello assunto in $\alpha = 6.495$, dove $\log L = 3.943$.

• Pozzo 14

Per questo pozzo, le funzioni utilizzate hanno stimato entrambe l'asimmetria con un valore positivo che tende a ∞ ; per questo si analizza il comportamento della funzione di log-verosimiglianza profilo relativa in intervalli centrati nei valori stimati dalle due funzioni.

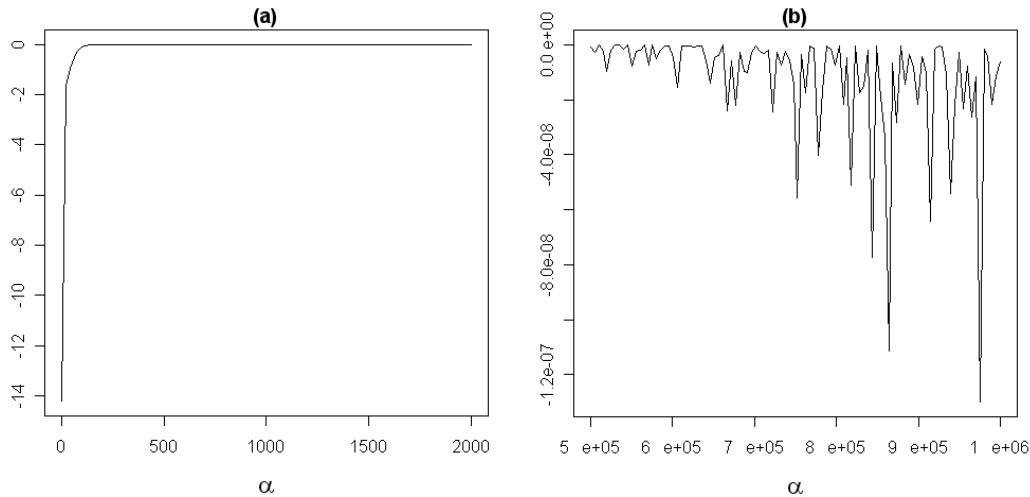


Figura 3.7 Pozzo 14: due volte la verosimiglianza profilo relativa di α in intervalli centrati nel valore stimato dalla funzione *sn.mle* (a) e *msn.mle* (b).

$0 \leq \alpha \leq 2000$		$500000 \leq \alpha \leq 1000000$	
α	$\log L$	α	$\log L$
2.02×10^2	-6.237	5.10×10^5	-6.239

Tabella 3.7 Pozzo 14: valore dei parametri e della log-verosimiglianza secondo i due intervalli considerati nei grafici (a) - (b).

La Figura 3.7 (a) analizza α nell'intervallo $0 \leq \alpha \leq 2'000$, dove viene segnalato un punto di massimo, differente da quello ottenuto tramite la funzione *sn.mle*, in prossimità di $\alpha = 202.020$ nel quale la log-verosimiglianza vale -6.237 . Nel grafico (b), invece, la presenza di oscillazione della funzione di log-verosimiglianza profilo relativa è più evidente. Nell'intervallo per cui $500'000 \leq \alpha \leq 1'000'000$,

viene individuato un punto di massimo locale in $\alpha = 510.501$, in cui $\log L = -6.239$. Nonostante questi risultati siano diversi da quelli ottenuti applicando i due diversi metodi di stima, le conclusioni sono le medesime: i valori relativi alla log-verosimiglianza in corrispondenza dei punti di massimo locale individuati non discostano in maniera rilevante l'uno dall'altro e non permettono di identificare in modo univoco il massimo assoluto ma la stima di α tende a ∞ .

• Pozzo 19

Per l'analisi di questo pozzo si segue lo stesso procedimento fatto per il pozzo 14.

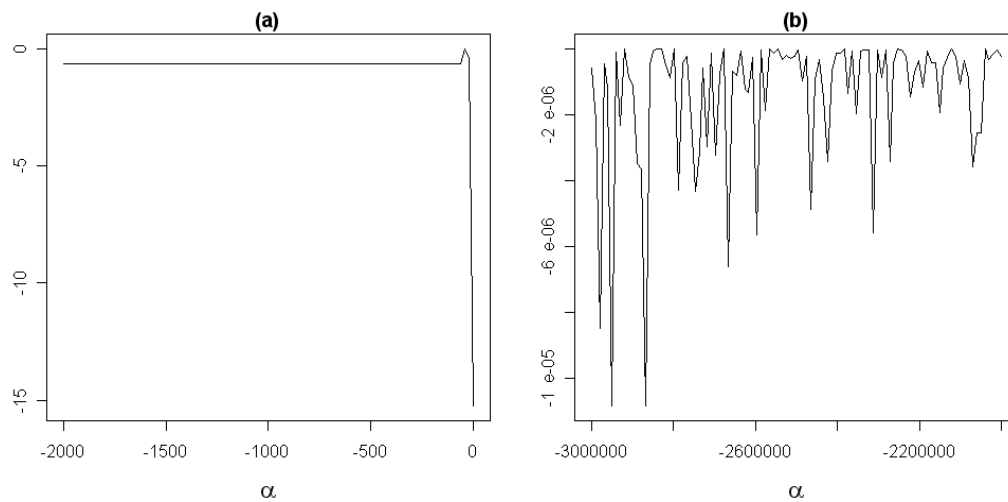


Figura 3.8 Pozzo 19: due volte la verosimiglianza profilo relativa di α in intervalli centrati nel valore stimato dalla funzione *sn.mle* (a) e *msn.mle* (b).

$-2000 \leq \alpha \leq 0$		$-3000000 \leq \alpha \leq -2000000$	
α	$\log L$	α	$\log L$
-40.404	-32.958	-2.91×10^6	-33.290

Tabella 3.8 Pozzo 19: valore dei parametri e della log-verosimiglianza secondo i due intervalli considerati nei grafici (a) - (b).

I metodi di stima utilizzati indicano la presenza di asimmetria negativa, ma in entrambi $\alpha \rightarrow -\infty$; per questo si analizza il comportamento della funzione di log-verosimiglianza profilo relativa in intervalli centrati nei valori stimati dalle due funzioni. La Figura 3.8 (a) analizza α nell'intervallo $-2000 \leq \alpha \leq 0$, dove viene segnalato un punto di massimo, differente da quello ottenuto tramite la funzione *sn.mle*, in prossimità di $\alpha = -40.404$ nel quale la log-verosimiglianza vale

-32.958.

Nel grafico (b) le oscillazioni della funzione di log-verosimiglianza profilo relativa sono molto evidenti. Nell'intervallo $-3'000'000 \leq \alpha \leq -2'000'000$ vi è un punto di massimo locale in $\alpha = -2.91 \times 10^6$, in cui $\log L = -33.290$. Nonostante questi risultati siano diversi da quelli ottenuti applicando i due diversi metodi di stima, le conclusioni sono le medesime: i valori relativi alla log-verosimiglianza in corrispondenza dei punti di massimo locale individuati non si discostano in maniera rilevante l'uno dall'altro e l'indice di asimmetria viene stimato con valori che tendono a $-\infty$.

Gli ultimi tre pozzi che seguono, presentano caratteristiche simili ai primi analizzati.

- **Pozzo 20**

I grafici (a) e (b) in Figura 3.9 concordano con i risultati ottenuti con la funzione *sn.mle* (la verosimiglianza profilo relativa vale zero nel punto di massimo che coincide con la stima ottenuta e, nelle curve di livello, il punto di coordinate $(\hat{\omega}, \hat{\alpha})$ è posizionato nella regione di confidenza più interna). Analizzando il riquadro (c), in cui α è valutato nell'intervallo $-20'000'000 \leq \alpha \leq -3$, si nota la presenza di oscillazioni che determinano più punti di massimo locale; questa situazione è confermata anche nel grafico (d) in cui si considera α nell'intervallo $-20'000'000 \leq \alpha \leq -15'000'000$ e vi è la presenza di un punto di massimo in $\alpha = 1.98 \times 10^7$ e la log-verosimiglianza vale -17.49. Questo risultato conferma che il massimo assoluto è il punto individuato dalla funzione *sn.mle*, in quanto la log-verosimiglianza in esso vale -12.497.

$-5 \leq \alpha \leq -2.5$			$-20'000'000 \leq \alpha \leq -15'000'000$	
ω	α	$\log L$	α	$\log L$
0.557	-3.475	-12.497	-1.98×10^7	-17.498

Tabella 3.9 Pozzo 20: valore dei parametri e della log-verosimiglianza secondo i due intervalli considerati nei grafici (a) - (d).

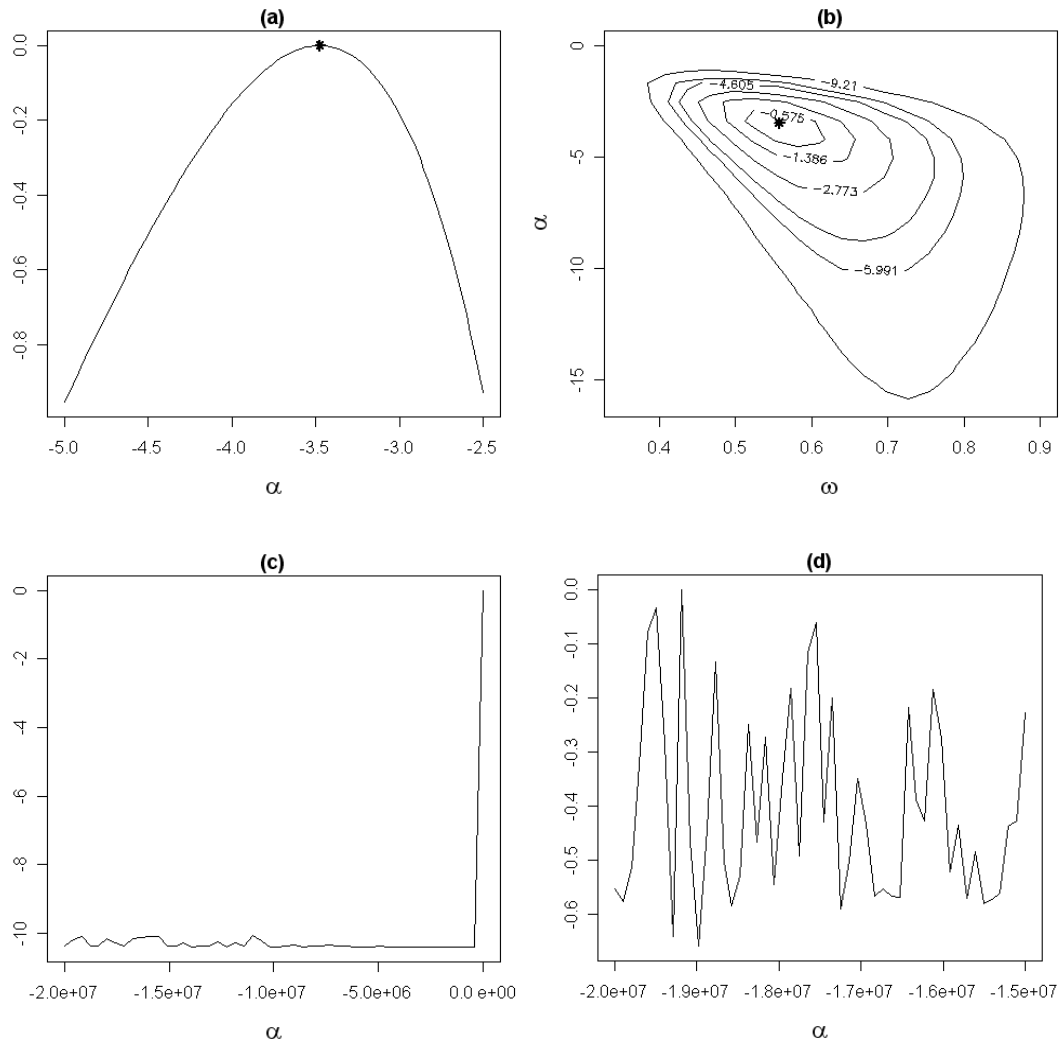


Figura 3.9 Pozzo 20: due volte la verosimiglianza profilo relativa di α e curve di livello della stessa funzione per (ω, α) verificando le stime ottenute tramite *sn.mle* (a-b) e due volte la log-verosimiglianza profilo relativa di α in intervalli più ampi (c-d).

• Pozzo 23

Nei grafici (a) e (b) le stime concordano con quelle ottenute tramite la funzione *sn.mle* (la verosimiglianza profilo relativa vale zero nel punto di massimo indicato dalla stima ottenuta e il punto di coordinate $(\hat{\omega}, \hat{\alpha})$ è posizionato nella regione di confidenza più interna delle curve di livello). Analizzando il riquadro (c) (α è valutato nell'intervallo $6 \leq \alpha \leq 750'000$) graficamente non è possibile individuare la presenza di punti di massimo locale; andando ad analizzarne il grafico (d), invece,

risulta un punto di massimo in $\alpha = 7.21 \times 10^5$ e in esso la log-verosimiglianza vale -28.408 ; questo valore conferma che il massimo assoluto è quello individuato dalla funzione *sn.mle* in cui $\log L = -22.727$.

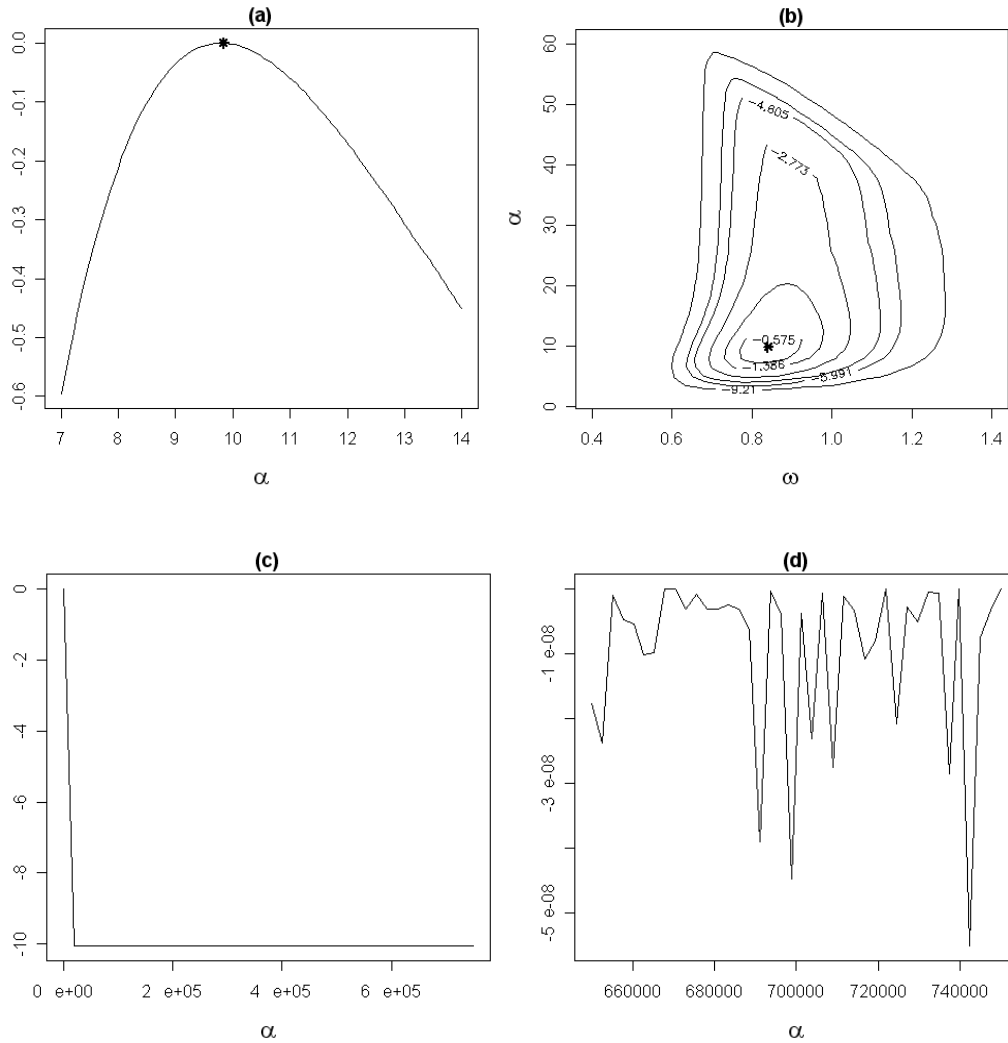


Figura 3.10 Pozzo 23: due volte la verosimiglianza profilo relativa di α e curve di livello della stessa funzione per (ω, α) verificando le stime ottenute tramite *sn.mle* (a-b) e due volte la log-verosimiglianza profilo relativa di α in intervalli più ampi (c-d).

<i>sn.mle</i>			<i>msn.mle</i>	
ω	α	$\log L$	α	$\log L$
0.840	9.837	-22.727	6.97×10^5	-28.408

Tabella 3.10 Pozzo 23: valore dei parametri e della log-verosimiglianza secondo i due intervalli considerati nei grafici (a) - (d).

• **Pozzo 24**

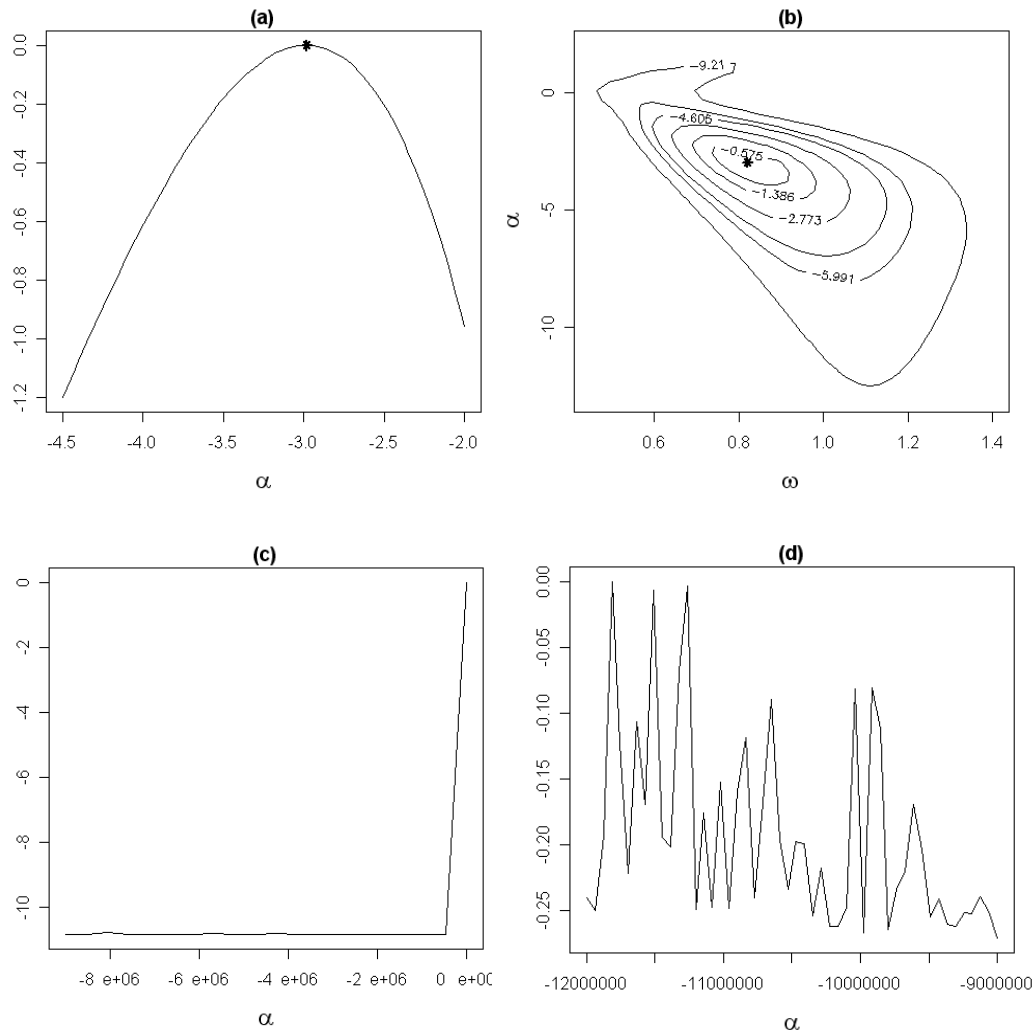


Figura 3.11 Pozzo 24: due volte la verosimiglianza profilo relativa di α e curve di livello della stessa funzione per (ω, α) verificando le stime ottenute tramite *sn.mle* (a-b) e due volte la log-verosimiglianza profilo relativa di α in intervalli più ampi (c-d).

<i>sn.mle</i>			<i>msn.mle</i>	
ω	α	$\log L$	α	$\log L$
0.820	-2.981	-27.868	-1.1×10^7	-33.626

Tabella 3.11 Pozzo 24: valore dei parametri e della log-verosimiglianza secondo i due intervalli considerati nei grafici (a) - (d).

Anche per quest'ultimo pozzo si conclude che la stima ottenuta con la funzione *sn.mle* indica l'effettivo punto di massimo assoluto: il valore della log-verosimiglianza in $\alpha = -2.981$ è -27.868 mentre nell'altro punto di massimo locale assume valore inferiore ($\log L = -33.626$).

Dai risultati ottenuti si ha avuto conferma della maggior affidabilità della funzione *sn.mle* in quanto la log-verosimiglianza nei punti di massimo risultanti è superiore e stima in maniera corretta i parametri della distribuzione scelta per spiegare i dati.

3.4 Stima del parametro di asimmetria tramite regressione locale

In questo paragrafo si esamina la possibile associazione tra la posizione del pozzo e il parametro di asimmetria della relativa distribuzione normale asimmetrica. La localizzazione del pozzo è rappresentata dalle coordinate (x_1, x_2) di un punto in un piano cartesiano (x_1, x_2) ⁴.

Per mettere in relazione le variabili si fa riferimento ad una formulazione del tipo $y = f(x_1, x_2) + \varepsilon$ dove ε è un termine di errore casuale non osservato con media zero e $f(x_1, x_2)$ è una funzione da \mathbb{R}^2 a \mathbb{R} . Tale formulazione ricorda quello della regressione lineare semplice ma in questo caso non si presume che f sia membro di una specificata classe parametrica.

Se $f(x_1, x_2)$ è una funzione differenziabile in $x_0 = (x_{01}, x_{02})$, punto fissato dei numeri reali, in base allo sviluppo in serie di Taylor è localmente approssimabile con una retta passante per $((x_{01}, x_{02}), f(x_{01}, x_{02}))$, cioè

$$f(x_1, x_2) = \underbrace{f(x_{01}, x_{02})}_{\alpha} + \underbrace{f'_{x_1}(x_{01}, x_{02})}_{\beta}(x_1 - x_{01}) + \underbrace{f'_{x_2}(x_{01}, x_{02})}_{\gamma}(x_2 - x_{02}) + \text{resto}$$

Per stimare $f(x_1, x_2)$ in un intorno di x_0 mediante un criterio che tenga conto di questo risultato sulla base di n coppie di osservazioni (x_i, y_i) per $i=1, \dots, n$, con $x_i = (x_{i1}, x_{i2})$, si utilizza il criterio dei minimi quadrati ma pesando le osservazioni

⁴ Nel paragrafo 1.2 è stata descritta la procedura della creazione del piano cartesiano e le relative coordinate che identificano ogni pozzo.

x_i in base alla loro distanza da x_0 :

$$\min_{\alpha, \beta, \gamma} \sum_{i=1}^n \{y_i - \alpha - \beta(x_{i1} - x_{01}) - \gamma(x_{i2} - x_{02})\}^2 w(x_{i1} - x_{01}; h_1) w(x_{i2} - x_{02}; h_2)$$

Questa è una forma particolare del criterio dei minimi quadrati pesati; visto che i pesi sono costruiti attorno a x_0 il metodo di stima risultante è detto regressione locale.

Il parametro h è un numero positivo che rappresenta un fattore di scala ed è detto parametro di lisciamento; $w(\cdot; h)$ è una funzione di densità simmetrica attorno all'origine, detta nucleo ed assicura che i pesi maggiori vengano assegnati alle osservazioni x_i più vicine al punto di interesse x_0 .

Il parametro di lisciamento controlla l'ampiezza del nucleo e quindi il grado di lisciamento applicato ai dati. Nelle analisi che seguono, verrà considerato il nucleo normale corrispondente alla densità $N(0,1)$; in tal caso, h è la deviazione standard. Considerando la Normale usuale e un livello di significatività del 5%, i valori significativi della distribuzione sono contenuti nell'intervallo delimitato dai quantili $(1 - \alpha/2)$ di $N(0, h^2)$ che corrisponde all'intervallo $[-1.96h, 1.96h]$. Le osservazioni contenute nell'intervallo di ampiezza $4h$ sull'asse della variabile contribuiranno, quindi, alla stima.

Disponendo di due variabili esplicative vengono coinvolti altrettanti parametri di lisciamento che, riferendo entrambi a coordinate cartesiane, non devono tenere conto della diversa variabilità di x_1 e x_2 , ma assumono il medesimo valore ($h = (h_1, h_2)$ dove $h_1 = h_2$).

Al crescere del parametro di lisciamento la stima risultante perde alcuni dettagli nella curva che interpola i dati e risulta più liscia; al contrario, al diminuire del parametro di lisciamento la stima risulta sempre più accurata fino ad ottenere l'interpolazione di tutti i punti osservati.

La Figura 3.12 riporta la rappresentazione grafica tramite curve di livello e grafico prospettico del risultato della stima del parametro di asimmetria dei pozzi tramite la regressione locale, usando come variabili esplicative x_1 = ascissa della posizione

del pozzo, x_2 =ordinata della posizione del pozzo. Per evitare di estrapolare la stima laddove non si disponga di osservazioni, questa viene limitata all'involucro convesso dei punti osservati delle coordinate (x_1, x_2) ⁵.

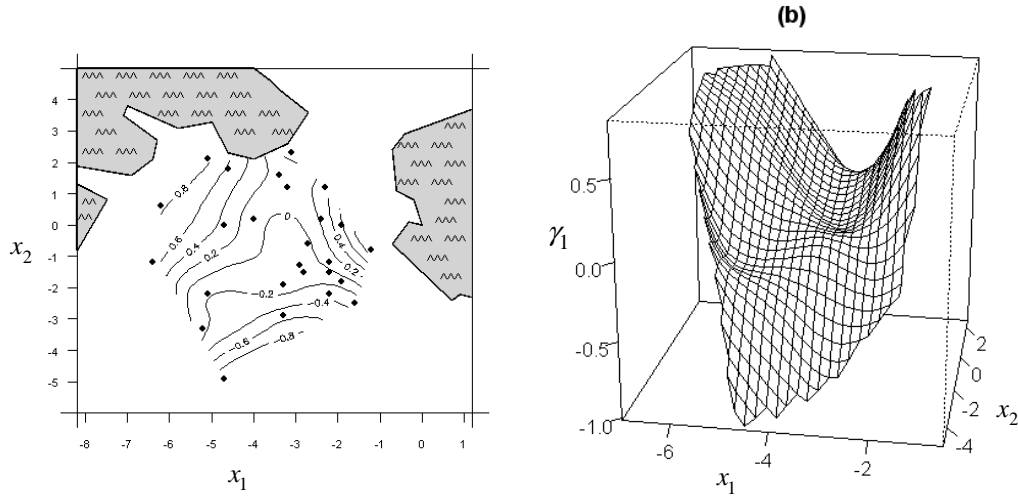


Figura 3.12 Curve di livello (a) e grafico prospettico (b) della stima del parametro di asimmetria (γ_1) tramite la regressione locale con due variabili esplicative espresse dalle coordinate dei punti che localizzano i pozzi ($h=0.898$).

Va precisato che per la creazione di tale immagine si è tenuto conto che nelle stime precedenti il parametro di asimmetria per alcuni pozzi è risultato $\alpha \rightarrow \infty$. Finora è stata considerata la distribuzione normale asimmetrica con parametri (ξ, ω, α) , $Y \sim \text{SN}(\xi, \omega^2, \alpha)$. Tali parametri, detti parametri diretti (Direct Parameters, DP), possono indurre a due problematiche: innanzitutto nella log-verosimiglianza profilo c'è sempre un punto di flesso in $\alpha=0$ che porta l'informazione attesa di Fisher a diventare singolare; inoltre la funzione di verosimiglianza può avere una forma che può essere lontana da quella quadratica quando α non è vicino a zero. Per risolvere la prima, Azzalini (1985) effettuò una riparametrizzazione utilizzando la trasformazione $Y = \mu + \sigma Z^0$, dove $Z^0 = (Z - \mu_z) / \sigma_z$, $\sigma_z = (1 - \mu_z^2)^{1/2}$, e si considerano i parametri centrati

⁵ Per la creazione dei grafici che seguono è stata utilizzata la funzione *sm.regression* contenuta nel pacchetto *sm* di R.

$CP=(\mu, \sigma, \gamma_1)$. In questo modo, i valori dell'indice di asimmetria γ_1 risultano in un intervallo chiuso per cui $|\gamma_1| \leq 0.99527$.

Dato che per alcuni pozzi la stima del parametro di asimmetria è risultata $\alpha \rightarrow \infty$, per ovviare a questo problema e ottenere rappresentazioni grafiche più appropriate, si è ricorsi all'utilizzo dei parametri centrati. La Tabella 3.12 riporta la stima del parametro di asimmetria α e il relativo valore riparametrizzato γ_1 .

Pozzo	α	γ_1	Pozzo	α	γ_1
1	1.3582	0.25303	14	1.5337×10^3	0.99527
2	-1.8095	-0.39863	15	-2.0258	-0.46094
3	-2.0763	-0.47460	16	2.0117	0.45706
4	7.0798	0.91859	17	-2.6816	-0.61214
5	2.1119	0.48403	18	0.9237	0.11460
6	3.2015	0.69664	19	-1.5337×10^3	-0.99527
7	4.0672	0.79012	20	-3.4750	-0.73149
8	-3.8517	-0.77107	21	3.5625	0.74147
9	0.4902	0.02265	22	-2.5265	-0.58135
10	6.9454	0.91580	23	9.8375	0.95428
11	1.1823	0.19457	24	-2.9818	-0.66416
12	-1.1852	-0.19552	25	1.4814	0.29411
13	0.1382	0.00057	26	0.0837	0.00013

Tabella 3.12 Valore dell'indice di asimmetria α e il corrispondente valore riparametrizzato γ_1 .

Dal grafico (a) in Figura 3.12 è possibile notare la presenza di una forma di associazione tra la posizione del pozzo e l'indice di asimmetria. Si possono infatti individuare fasce distinte in base al segno di γ_1 : partendo dal basso (zona più vicina al vulcano) si trovano curve di livello orizzontali con stime negative che, in valore assoluto, decrescono fino alla parte centrale della zona considerata; da qui le linee hanno inclinazione verticale e le stime cambiano segno diventando positive e assumendo valori crescenti man mano ci si sposta dal centro verso la costa del Porto di Levante o del Porto di Ponente. Queste considerazioni vengono confermate dal grafico prospettico (b)⁶, in cui la parte superiore (dove l'asse delle ordinate x_2 è positiva) presenta forma parabolica con concavità verso l'alto e il vertice rappresenta la fascia centrale della zona esaminata le cui stime sono molto vicine a zero; le parti esterne più alte rappresentano la vicinanza alle coste del Porto di Levante

⁶ Per capire i valori a cui fanno riferimento gli assi riportati, consultare il paragrafo 1.2 in cui viene descritta la procedura per la costruzione del piano cartesiano di riferimento.

e del Porto di Ponente in cui i valori di γ_1 sono positivi e crescono allontanandosi sempre più dalla fascia centrale. Considerando invece la parte inferiore del grafico (asse delle ordinate x_2 con valore inferiore a zero) le stime hanno segno negativo e non variano in base alla vicinanza del pozzo alle coste ma assumono valori inferiori in prossimità del vulcano.

Come spiegato in precedenza, la scelta del valore di h stabilisce il livello di lisciamento della curva stimata. Per la creazione del grafico sopra riportato, tale valore è stato calcolato dal programma ed è risultato pari a $h_1 = h_2 = 0.898$; le rappresentazioni che seguono, invece, aiutano a comprendere le differenze che subiscono le stime per altre scelte di h .

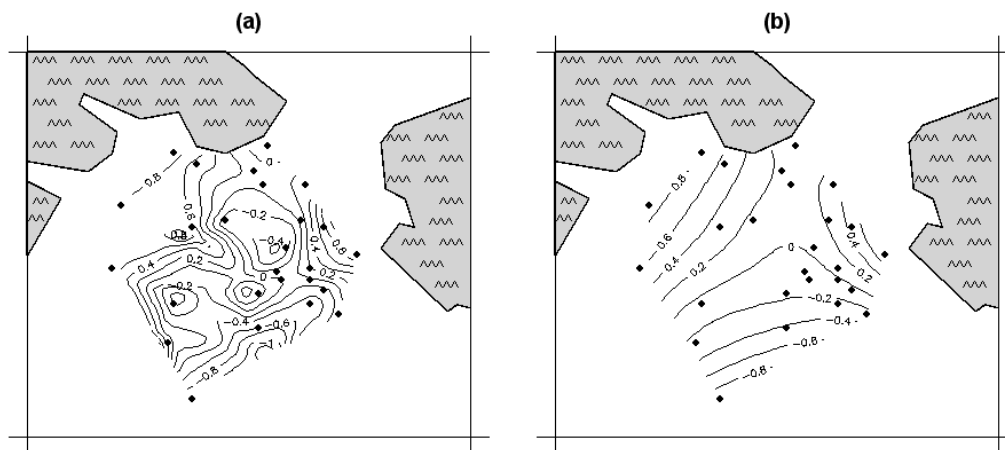


Figura 3.13 Curve di livello della stima del parametro di asimmetria (γ_1) tramite la regressione locale con due variabili esplicative espresse dalle coordinate dei punti che localizzano i pozzi, fissando $h=0.5$ (a) e $h=1.2$ (b).

Nei grafici in Figura 3.13 si riconosce lo stesso andamento di fondo riscontrato in Figura 3.12 nonostante in (b) le curve presentino minori fluttuazioni dovute all'aumento del valore di h e in (a) la stima risulti più accurata (h più basso). In quest'ultimo pannello, a differenza degli altri, le curve sono più aderenti al comportamento dei dati ed è possibile individuare più punti di massimo nella parte centrale della zona considerata, da cui le curve di livello si estendono verso

l'esterno indicando stime positive decrescenti nella zona superiore e in quella inferiore stime negative in valore assoluto crescenti.

Si conclude che il parametro di asimmetria risulta fattore discriminante per i pozzi; infatti, note le coordinate che li localizzano, si può conoscere il segno dell'indice di forma e individuare se l'asimmetria si presenta in maniera più o meno accentuata nella distribuzione di ogni pozzo.

3.5 Stima della mediana tramite regressione locale

Il procedimento appena esaminato viene ripetuto per verificare l'esistenza di una relazione tra le osservazioni registrate per ogni pozzo e la localizzazione di ognuno di essi.

La variabile analizzata è stata sintetizzata da un unico valore per ogni pozzo tramite la mediana; si è scelto l'utilizzo di questo indice di posizione in quanto, a differenza della media, non risente della presenza delle osservazioni anomale riscontrate durante l'analisi preliminare nel paragrafo 2.2.

I valori delle mediane riportate nella tabella soprastante indicano che tale indice assume solo valori negativi; questo è un risultato che non desta preoccupazioni: come riscontrato durante l'analisi preliminare, il dataset contiene solo 18 osservazioni positive.

Pozzo	Mediana	Pozzo	Mediana
1	-0.5247	14	-0.7205
2	-0.8050	15	-0.9601
3	-0.5897	16	-1.2148
4	-1.0912	17	-2.2727
5	-1.1019	18	-1.5433
6	-0.9738	19	-1.4434
7	-0.9332	20	-1.2183
8	-0.7420	21	-1.5593
9	-1.7343	22	-1.0138
10	-0.8665	23	-0.6967
11	-1.7919	24	-2.5263
12	-1.7900	25	-1.1905
13	-0.9678	26	-0.9657

Tabella 3.13 Mediana delle osservazioni registrate per ogni pozzo.

In maniera analoga all'indice di asimmetria, non è possibile riconoscere nelle curve di livello delle stime della mediana andamento monotono. Come riportato in Figura 3.14 (a), partendo dalla parte inferiore del grafico, le stime in valore assoluto decrescono fino alla fascia centrale, punto da cui crescono nuovamente man mano ci si avvicina alla costa del Porto di Ponente. Questa situazione è confermata dal grafico in (b) in cui le stime risultanti presentano forma parabolica con concavità verso il basso e il vertice è in corrispondenza alla parte centrale della zona esaminata.

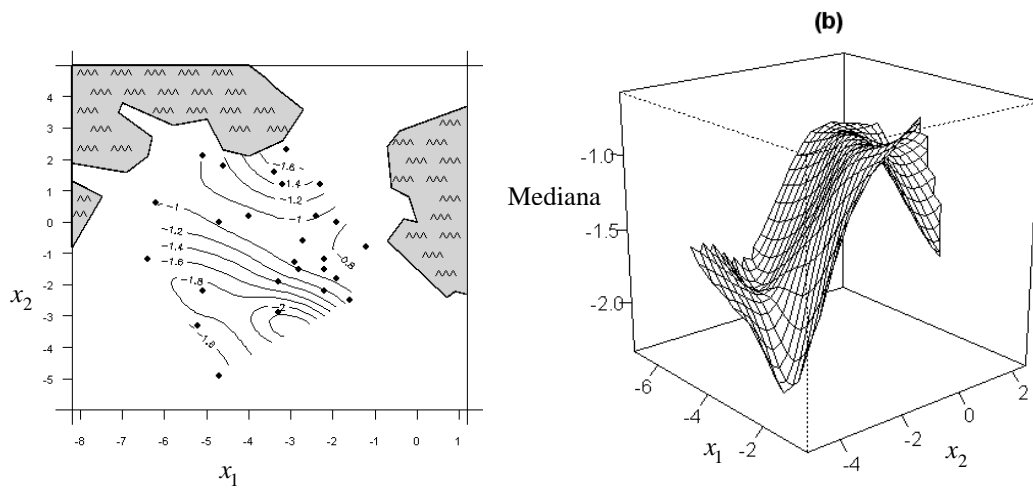


Figura 3.14 Curve di livello (a) e grafico prospettico (b) della stima della mediana tramite la regressione locale con due variabili esplicative espresse dalle coordinate dei punti che localizzano i pozzi.

Confrontando questi risultati con quelli ottenuti nel caso precedente, le stime mostrano somiglianze nella zona più vicina al vulcano, mentre la parte superiore si presenta in maniera opposta: nel primo caso ha forma parabolica con concavità verso l'alto e la fascia centrale rappresenta un minimo; le stime della mediana, invece, sono rappresentate graficamente con andamento parabolico con concavità verso il basso e le coste rappresentano i minimi.

Si esaminano ora le modifiche che subiscono le curve di livello al variare del parametro h , che stabilisce il livello di lisciamento della curva stimata. Come per le stime dell'indice di asimmetria, per il grafico in Figura 3.14 (a) tale valore è stato

fissato dal programma pari a $h_1 = h_2 = 0.898$; le rappresentazioni in Figura 3.15 riportano le differenze che si registrano diminuendo ($h=0.5$) e aumentando ($h=1.2$) il valore assegnato al parametro di lisciamento.

Per meglio valutare i grafici, le stime risultate negative, verranno trattate in valore assoluto.

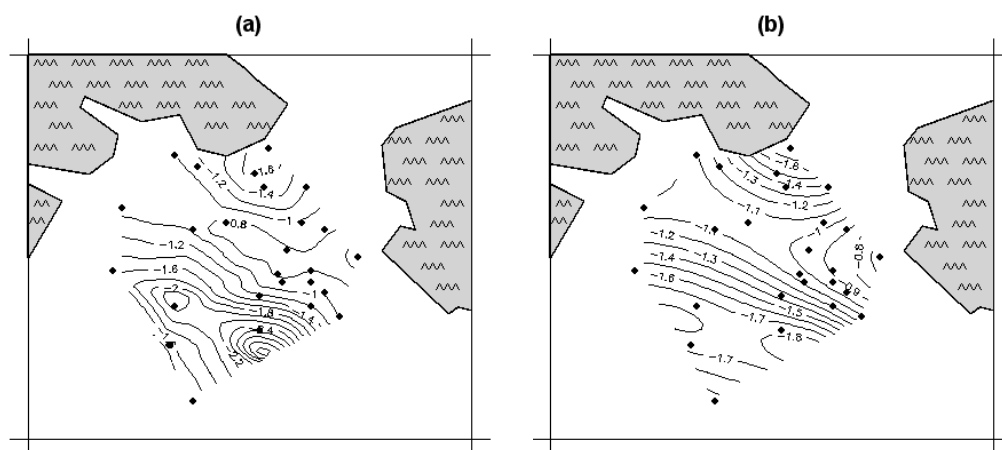


Figura 3.15 Curve di livello della stima della mediana tramite la regressione locale con due variabili esplicative espresse dalle coordinate dei punti che localizzano i pozzi, fissando $h=0.5$ (a) e $h=1.2$ (b).

Nel pannello di sinistra (valore di h più piccolo) si riconosce lo stesso comportamento riscontrato per $h=0.898$, ma è possibile identificare due punti di massimo nella parte inferiore del grafico da cui le stime decrescono spostandosi verso il vulcano e verso l'alto. La fascia centrale rappresenta ancora la zona con stime più basse, ma a differenza delle considerazioni fatte in precedenza, in essa vengono raggruppati i pozzi rivolti verso la costa del Porto di Levante. Spostandosi invece verso la costa del Porto di Ponente le curve di livello indicano stime con valori crescenti.

Le stesse considerazioni vengono fatte anche per il pannello di destra, anche se i due massimi non risultano ben distinti come in quello di sinistra.

(Si ricorda che queste conclusioni riferiscono alle stime considerate in valore assoluto).

Dai risultati ottenuti si può affermare che l'asse delle ordinate, indicatore della vicinanza dei pozzi al vulcano, è un elemento importante per capire per ognuno di essi la natura dell'asimmetria della distribuzione della variabile e i valori assunti da questa nel descrivere il rapporto tra catione calcio e anione solfato.

3.6 Stima della mediana tramite verosimiglianza locale

Nel paragrafo precedente è stata analizzata la stima della mediana tramite il metodo della regressione locale; l'utilizzo di questo metodo, però, ignora la natura della variabile considerata in quanto non richiede la specificazione di nessuna classe parametrica. Per rimediare a questa problematica e tenere conto della distribuzione da cui provengono i dati, è necessario applicare una generalizzazione al procedimento utilizzato: un'estensione naturale al meccanismo della regressione locale è la verosimiglianza locale ottenuta tramite l'applicazione di pesi alla log-verosimiglianza nella forma

$$l_{[h_1, h_2, x_{01}, x_{02}]}(\vartheta) = \sum_{i=1}^m \left[\sum_{j=1}^{n_j} l_{ij}(\vartheta) \right] w(x_{i1} - x_{01}; h_1) w(x_{i2} - x_{02}; h_2)$$

dove $l_{ij}(\vartheta)$ è il contributo della j -esima osservazione dell' i -esimo pozzo all'usuale log-verosimiglianza della distribuzione normale asimmetrica, $\vartheta = (\xi, \omega, \alpha)$, $m=26$ indica il numero di pozzi esaminati e n_j è la numerosità interna di ogni pozzo.

L'obiettivo è stimare il valore della mediana, indice di posizione scelto per sintetizzare i valori assunti dalla variabile, per un pozzo posizionato nel punto $x_0 = (x_{01}, x_{02})$ tenendo conto sia delle osservazioni registrate per i ventisei pozzi esaminati che della loro posizione.

È possibile notare che nella formula citata $w(\cdot; h)$ non dipende dall'indice j ; i pesi, infatti, sono determinati in base alla distanza tra la posizione del pozzo x_i e il punto di stima x_0 e non vi è l'influenza delle osservazioni registrate per la varia-

bile d'interesse.

La formula precedente può essere così riassunta

$$l_{[h_1, h_2, x_{01}, x_{02}]}(\vartheta) = \sum_{i=1}^m l_i(\vartheta) w(x_{i1} - x_{01}; h_1) w(x_{i2} - x_{02}; h_2)$$

Massimizzando la log-verosimiglianza rispetto a (ξ, ω, α) si ottengono le stime dei parametri della distribuzione normale asimmetrica del pozzo di coordinate x_0 .

Dopo aver creato una funzione per il calcolo della verosimiglianza pesata, si confrontano i valori dei parametri stimati per ogni pozzo con questo procedimento e tramite la funzione *sn.mle* utilizzata in precedenza nel paragrafo 3.2.

Per verificare se i pesi aggiunti generano effettivamente alcune differenze si deve annullare l'effetto del parametro di lisciamento assegnando a h un valore molto vicino a zero ($h=0.1$).

Pozzo	ξ	ω	α
1	-0.337	0.387	-0.849
2	-0.416	0.609	-1.761
3	-0.358	0.353	-2.076
4	-1.508	0.773	7.079
5	-1.585	0.766	2.112
6	-1.220	0.365	3.202
7	-1.337	0.599	4.067
8	-0.169	0.855	-3.852
9	-1.879	0.416	0.490
10	-1.118	0.399	6.945
11	-2.033	0.413	1.182
12	-1.474	0.550	-1.185
13	-0.985	0.140	0.158
14	-1.110	0.559	11.593
15	-0.821	0.231	-1.428
16	-1.448	0.355	2.012
17	-1.989	0.433	-2.682
18	-1.709	0.317	0.924
19	-0.705	1.055	-1533.768
20	-0.884	0.557	-3.475
21	-1.798	0.355	3.563
22	-0.896	0.176	-2.526
23	-1.132	0.788	1533.768
24	-2.039	0.821	-2.982
25	-1.509	0.512	1.481
26	-0.990	0.234	0.133

Tabella 3.14 Parametri stimati tramite verosimiglianza locale fissando $h=0.1$.

Pozzo	ξ	ω	α
1	-0.763	0.397	1.358
2	-0.440	0.612	-1.761
3	-0.358	0.353	-2.076
4	-1.508	0.773	7.079
5	-1.585	0.766	2.112
6	-1.220	0.365	3.202
7	-1.337	0.599	4.067
8	-0.169	0.855	-3.852
9	-1.879	0.416	0.490
10	-1.118	0.399	6.945
11	-2.033	0.413	1.182
12	-1.474	0.550	-1.185
13	-0.983	0.139	0.138
14	-1.113	0.556	1533.768
15	-0.803	0.241	-2.025
16	-1.448	0.355	2.012
17	-1.989	0.433	-2.682
18	-1.709	0.317	0.924
19	-0.703	1.096	-1533.768
20	-0.884	0.557	-3.475
21	-1.798	0.355	3.563
22	-0.896	0.176	-2.526
23	-1.054	0.841	9.837
24	-2.039	0.821	-2.982
25	-1.509	0.512	1.481
26	-0.981	0.233	8.374

Tabella 3.15 Parametri stimati tramite la funzione *sn.mle*.

Per alcuni pozzi le stime risultano le medesime, altre presentano differenze.

Considerando la cartina in Figura 1.2 che riporta la localizzazione dei ventisei pozzi, è possibile notare che per i pozzi isolati le stime ottenute tramite i due metodi non presentano differenze. In questi casi, infatti, i pesi sono tutti nulli tranne in corrispondenza al pozzo stesso dove x_i coincide con x_0 e solo la log-verosimiglianza di questo pozzo contribuisce alla log-verosimiglianza finale. Al contrario, per pozzi molto vicini, più pesi risultano non nulli e la log-verosimiglianza pesata è ottenuta dalla somma di più componenti diversi da zero.

Dopo aver constatato l'esistenza di differenze tramite l'utilizzo della verosimiglianza locale, si vogliono ora confrontare le diversità delle stime dei parametri, e di conseguenza della mediana, ottenute tramite la regressione locale e tramite la verosimiglianza locale secondo diversi valori del parametro di liscio h . Innanzitutto l'analisi non viene limitata ai punti osservati ma si estende il caso ad una griglia posizionata in un intervallo che ricopre i ventisei punti osservati; successivamente, tramite la verosimiglianza locale, sono stati stimati i parametri della distribuzione per ogni punto di intersezione nella griglia⁷. Infine, per la rappresentazione grafica, sono state disegnate le curve di livello della mediana relativa alle stime calcolate secondo diversi livelli di liscio. Per evitare di estrapolare la stima laddove non si dispone di osservazioni (la griglia si estende nell'acqua e in zone troppo lontane dai punti noti), questa viene rappresentata seguendo l'involucro convesso dei punti osservati. Questa restrizione è stata imposta anche per un altro motivo: se si fosse tenuto conto dell'intera griglia si sarebbe dovuto forzare la regressione locale ad estendere la stima nell'intero intervallo indicato portando a possibili alterazioni dei risultati e a confronti errati.

Di seguito vengono messi a confronto i grafici delle curve di livello delle stime della mediana tramite verosimiglianza locale e regressione locale secondo diversi valori del parametro di liscio.

⁷ La griglia è stata posizionata nell'intervallo $x=(-7,0.5)$, $y=(-6,3)$ con fittezza 20x20 quindi i punti di intersezione nella griglia risultano 400.

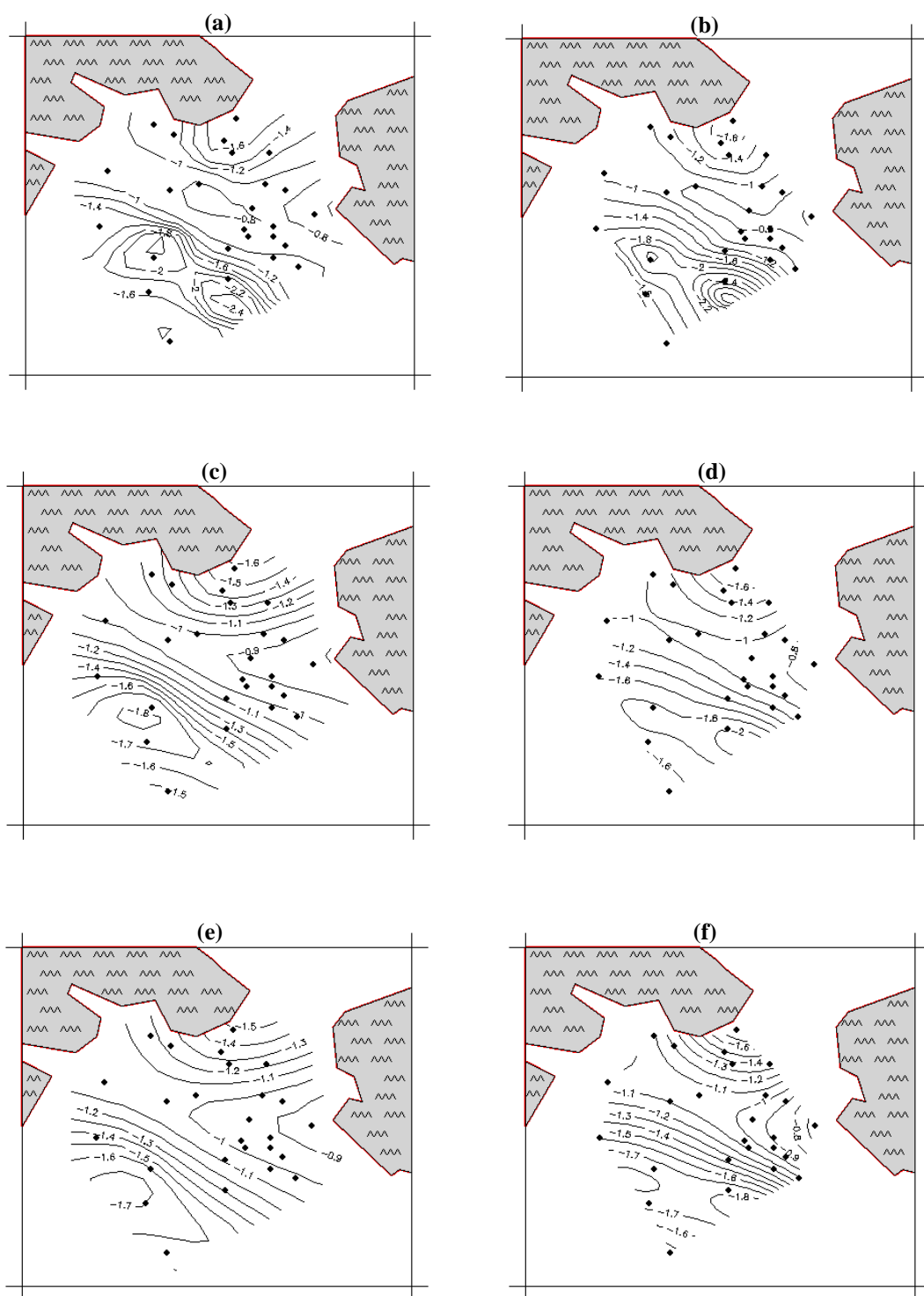


Figura 3.16 (a)-(f) Curve di livello della stima della mediana tramite la verosimiglianza locale (grafico di sinistra) e regressione locale (grafico di destra) con due variabili esplicative espresse dalle coordinate dei punti che localizzano i pozzi, fissando $h=0.6$ (a-b), $h=1$ (c-d), $h=1.2$ (e-f).

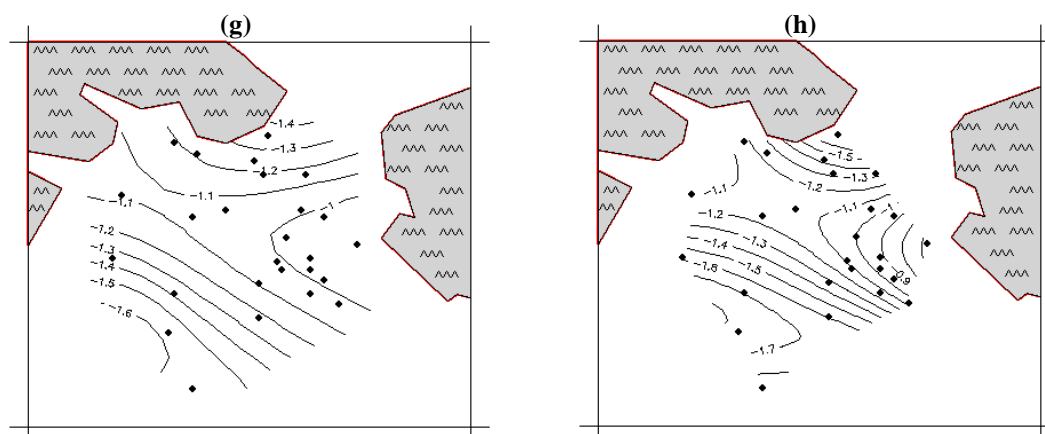


Figura 3.16 (g-h) Curve di livello della stima della mediana tramite la verosimiglianza locale (g) e regressione locale (h) con due variabili esplicative espresse dalle coordinate dei punti che localizzano i pozzi, fissando $h=1.5$.

In tutti i grafici è possibile suddividere la zona considerata in tre parti (inferiore, centrale e superiore) a seconda dell'andamento delle curve; la parte superiore, simile in tutti i riquadri presenta curve con valori negativi che in valore assoluto crescono avvicinandosi alla zona del Vulcanello. Mentre la parte centrale ha lo stesso andamento in entrambi i metodi di stima, quella inferiore presenta maggiori scostamenti. In particolare, nei pannelli di partenza in cui è stato posto $h=0.6$, entrambi i metodi di stima individuano due punti di massimo: uno verso la costa occidentale e uno più spostato verso il vulcano. Assegnando valore più alto al parametro di lisciamento ($h=1$), però, si presentano situazioni differenti: la verosimiglianza locale mantiene in maggiore evidenza il punto di massimo posizionato verso la costa occidentale e la regressione locale quello verso il vulcano. Per $h=1.2$, invece, la regressione locale individua nuovamente due punti di massimo e, aumentando ancora il livello di lisciamento, la situazione tende a coincidere con quella descritta dalla verosimiglianza locale.

Queste incongruenze trovano spiegazione nella posizione dei punti osservati: la parte inferiore occidentale è quella in cui vi è un numero inferiore di osservazioni note e con concentrazione molto più rada rispetto alle altre.

Capitolo 4

LA DISTRIBUZIONE t

ASIMMETRICA

Durante l'analisi preliminare nel paragrafo 2.2 è emerso che i dati presentano distribuzione con caratteristiche simili alla normale, ma con l'aggiunta di un parametro che ne regola la forma. Fino ad ora è stata considerata la distribuzione normale asimmetrica, ma un'altra distribuzione collegata a questa e che soddisfa i requisiti richiesti è la t asimmetrica, versione modificata della t di Student.

L'usuale costruzione della distribuzione t avviene tramite il rapporto tra una variabile normale e una chi-quadrato opportunamente trasformata. Se si vuole introdurre una variante alla distribuzione t che regola l'asimmetria, si sostituisce la variabile normale con la normale asimmetrica.

La distribuzione t asimmetrica corrisponde alla trasformazione

$$Y = \xi + V^{-1/2}Z, \text{ indicata con } Y \sim \text{St}(\xi, \omega, \alpha, \nu)$$

dove Z ha funzione densità normale asimmetrica $\text{SN}(0, \omega, \alpha)$ e $V \sim \chi^2_\nu / \nu$, indipendente da Z (Azzalini, A. e Capitanio, A., 2002).

Quando il parametro di forma α risulta pari a zero l'asimmetria svanisce e si riottiene la distribuzione t di Student, $Y \sim t(\xi, \omega, \nu)$; per ν sufficientemente elevato, invece, la distribuzione converge alla normale asimmetrica $Y \sim \text{SN}(\xi, \omega, \alpha)$.

4.1 Stima della distribuzione t asimmetrica

Per stimare i parametri della distribuzione t asimmetrica di ogni pozzo si segue la stessa procedura utilizzata nel paragrafo 3.2. Per tale scopo, il programma R prevede due funzioni, contenute anch'esse all'interno del pacchetto *sn* che, tramite la stima di massima verosimiglianza, permettono di ottenere una stima dei parametri $(\xi, \omega, \alpha, \nu)$ della distribuzione. La prima è la funzione *st.mle*, utilizzabile nei casi in cui la variabile osservata si presenti sotto forma di vettore; l'altra è la funzione *mst.mle* che estende l'analisi al caso multidimensionale (la variabile osservata può essere vettore o matrice). In maniera analoga alle considerazioni fatte per la normale asimmetrica, per stimare i parametri della distribuzione di ogni pozzo possono essere utilizzate entrambe le funzioni che, nonostante eseguano operazioni diverse, ci si aspetta diano le medesime stime⁸. Per verificare tale situazione, per ogni pozzo vengono riportate le curve stimate tramite le due funzioni (la linea tratteggiata indica la stima ottenuta con la funzione *st.mle*, la linea continua indica la stima ottenuta con la funzione *mst.mle*) e i valori dei parametri della distribuzione.

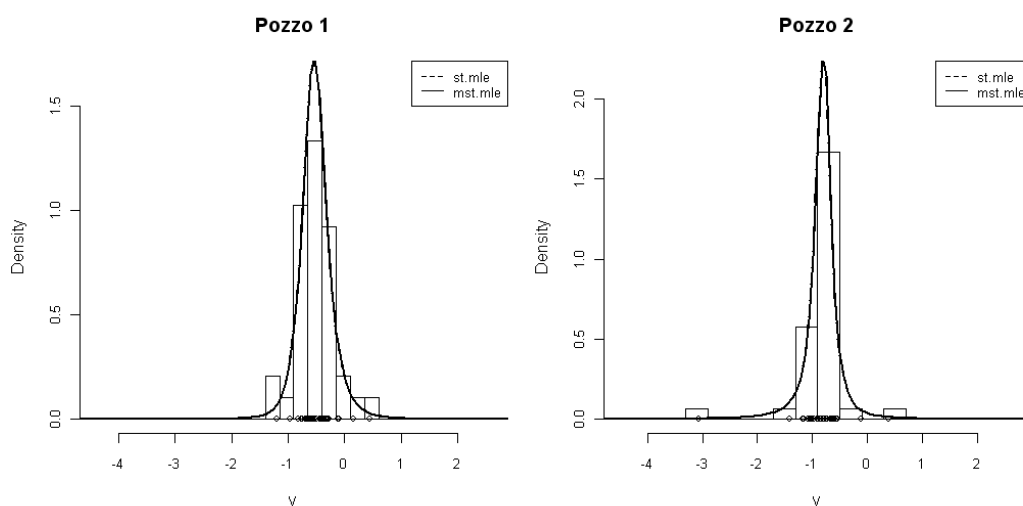


Figura 4.1 Pozzi 1-2 Istogramma con le curve della densità normale asimmetrica stimate tramite *st.mle* (linea tratteggiata) e tramite *mst.mle* (linea continua).

⁸ L'oggetto di studio è un vettore che descrive il logaritmo naturale del rapporto tra catione calcio e anione solfato e viene indicato nei grafici che seguono tramite la variabile y .

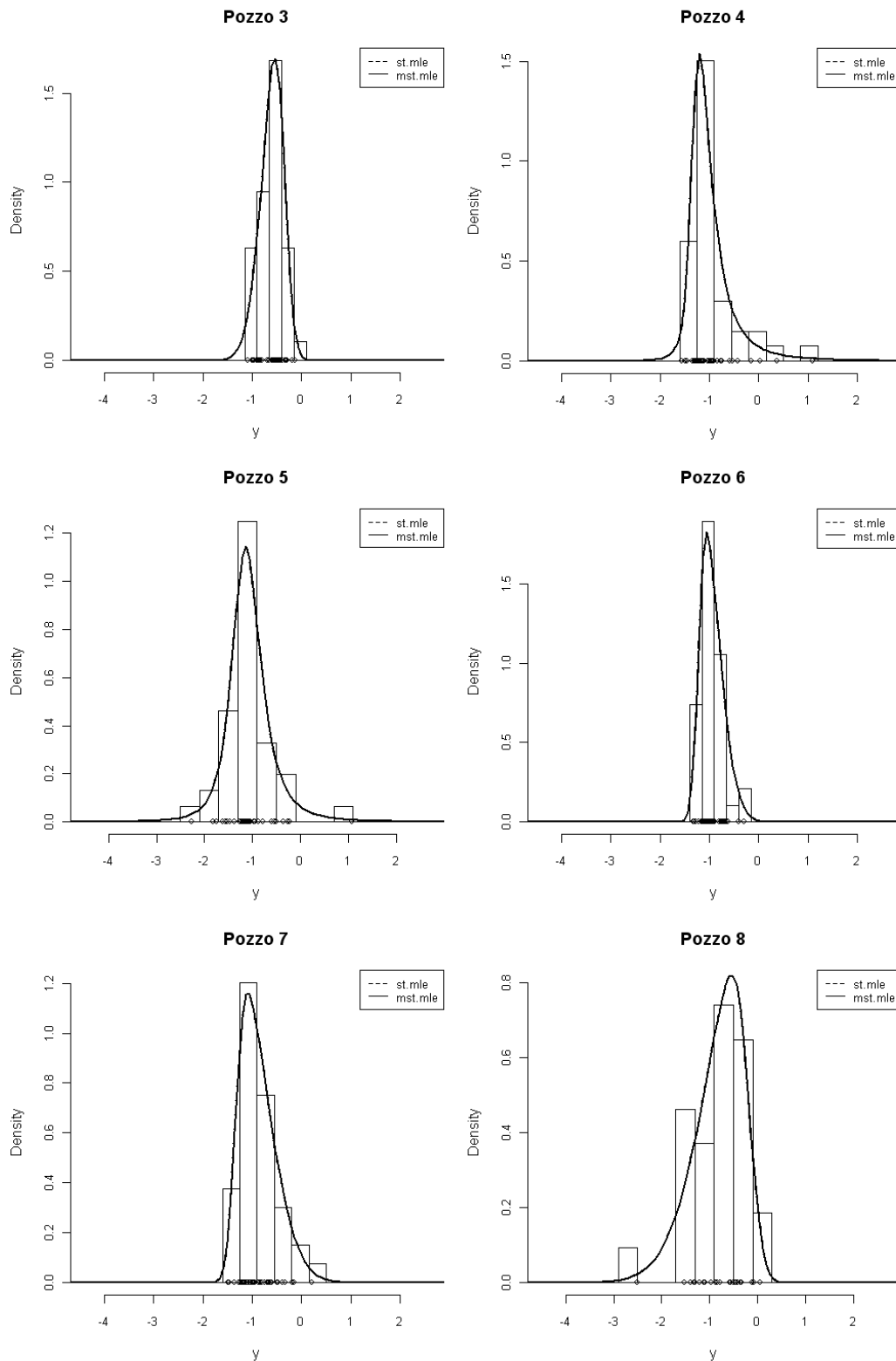


Figura 4.1 Pozzi 3-8 Istogramma con le curve della densità normale asimmetrica stimate tramite *st.mle* (linea tratteggiata) e tramite *mst.mle* (linea continua).

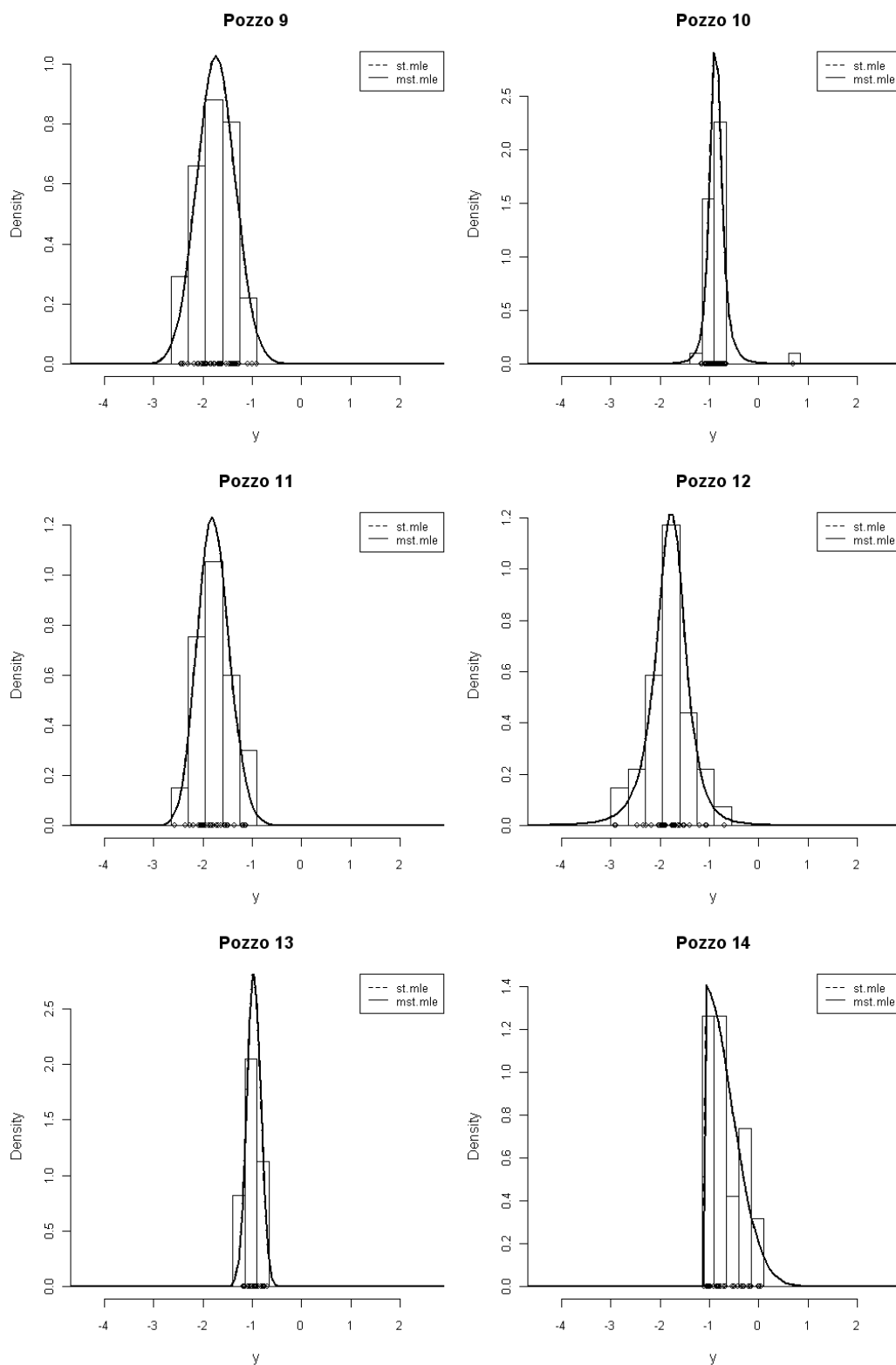


Figura 4.1 Pozzi 9-14 Istogramma con le curve della densità normale asimmetrica stimate tramite *st.mle* (linea tratteggiata) e tramite *mst.mle* (linea continua).

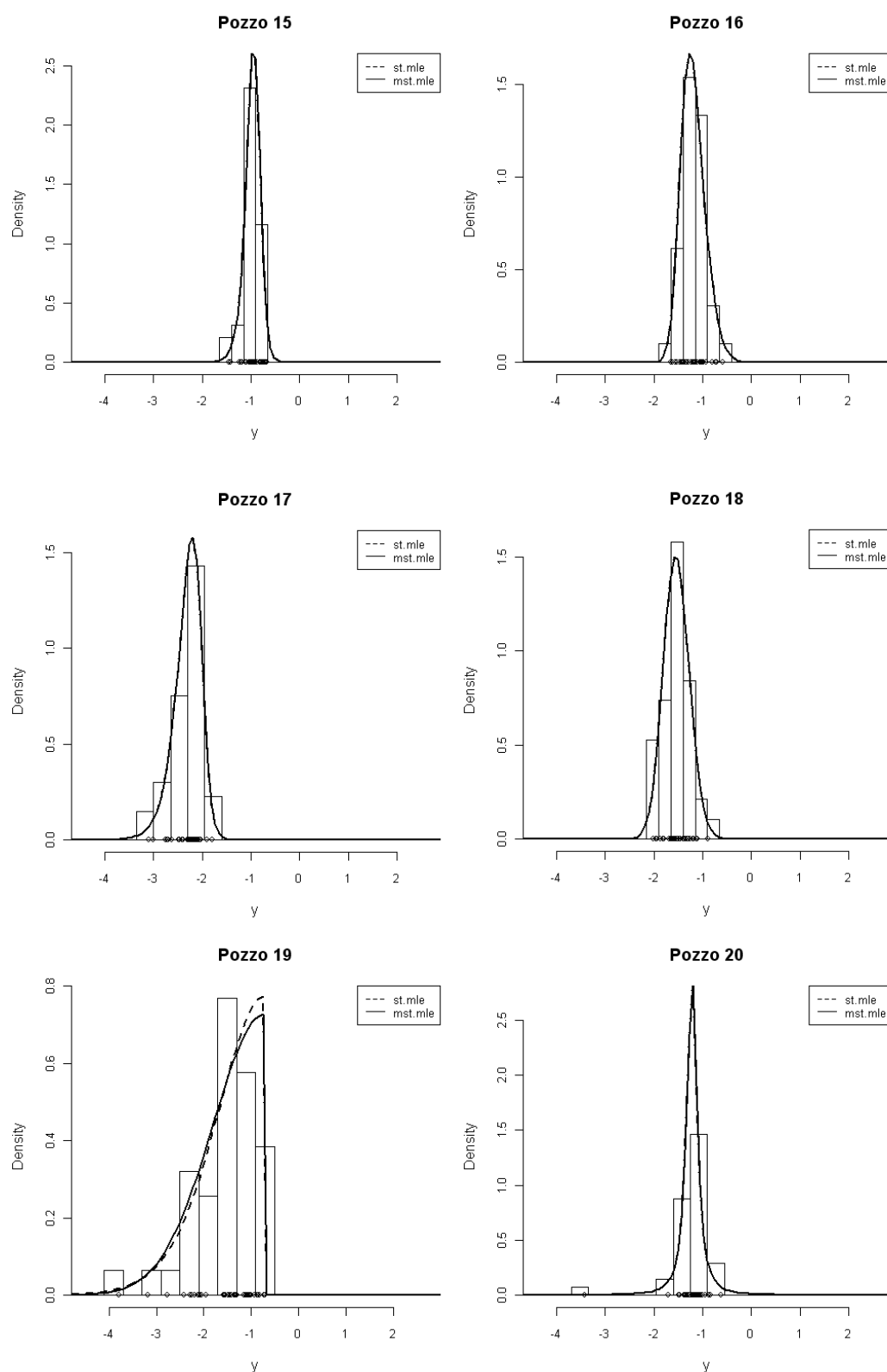


Figura 4.1 Pozzi 15-20 Istogramma con le curve della densità normale asimmetrica stimate tramite *st.mle* (linea tratteggiata) e tramite *mst.mle* (linea continua).

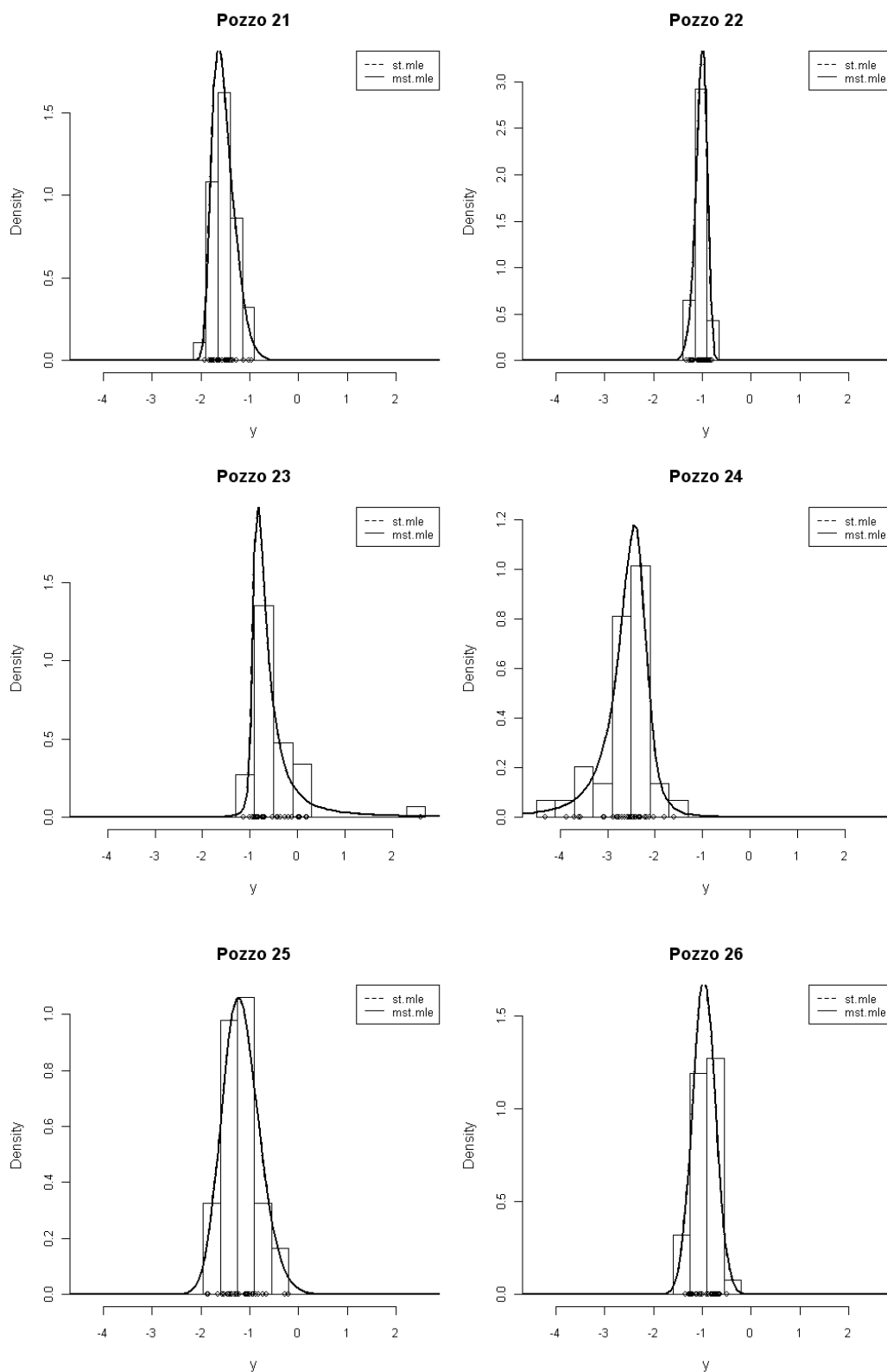


Figura 4.1 Pozzi 21-26 Istogramma con le curve della densità normale asimmetrica stimate tramite *st.mle* (linea tratteggiata) e tramite *mst.mle* (linea continua).

È immediato notare che nei grafici, a differenza della stima della distribuzione normale asimmetrica, solo per il pozzo 19 è visibilmente riscontrabile la differenza tra i due metodi di stima. Per verificare tale situazione si confrontano, ora, i valori delle stime ottenute.

Nella Tabella 4.1 vengono riportati il valore della log-verosimiglianza nel punto di massimo e i valori dei parametri stimati tramite la funzione *st.mle* supponendo che ogni pozzo Y ha distribuzione $Y \sim \text{St}(\xi, \omega, \alpha, \nu)$, con ξ, ω, α e ν rispettivamente parametri di posizione, di scala, di forma e gradi di libertà.

Pozzo	ξ	ω	α	ν	logL
1	-0.5904	0.2223	3.698×10^{-1}	3.516	-7.0130
2	-0.7703	0.1508	-2.588×10^{-1}	1.672	-6.3447
3	-0.3575	0.3528	-2.078	1.278×10^4	0.1398
4	-1.3412	0.3281	2.050	2.074	-18.0886
5	-1.2260	0.3350	4.705×10^{-1}	2.598	-25.9579
6	-1.2196	0.3654	3.201	9.190×10^3	2.5410
7	-1.3372	0.5991	4.067	1.177×10^4	-14.6441
8	-0.1821	0.8244	-3.593	4.443×10^1	-20.2216
9	-1.9009	0.4242	5.724×10^{-1}	3.670×10^5	-18.5587
10	-0.8941	0.1245	2.651×10^{-1}	2.917	12.4089
11	-2.0327	0.4132	1.183	9.014×10^3	-11.4280
12	-1.7114	0.3095	-3.098×10^{-1}	3.199	-21.5352
13	-0.9828	0.1393	1.363×10^{-1}	5.401×10^3	21.7485
14	-1.1134	0.5639	7.922×10^2	∞	-5.8366
15	-0.8537	0.1792	-1.136	7.857	14.6717
16	-1.4478	0.3553	2.010	1.210×10^4	-0.3895
17	-2.0235	0.3663	-2.136	9.946	-5.1950
18	-1.7092	0.3164	9.221×10^{-1}	1.001×10^4	-3.5714
19	-0.7064	1.0120	-1.606×10^6	1.336×10^1	-31.7425
20	-1.2173	0.1184	-9.036×10^{-3}	1.519	-0.2468
21	-1.7982	0.3545	3.560	1.602×10^4	4.2742
22	-0.8960	0.1764	-2.526	7.818×10^3	27.6155
23	-0.9297	0.2695	3.713	1.432	-15.1130
24	-2.2628	0.3842	-1.335	2.218	-26.2200
25	-1.5090	0.5117	1.481	1.399×10^4	-15.8815
26	-0.9954	0.2347	1.602×10^{-1}	1.852×10^5	1.3837

Tabella 4.1 Stime dei parametri ottenute tramite la funzione *st.mle* e valore della log-verosimiglianza nel massimo.

Dai valori riportati è possibile notare che per dieci pozzi il parametro di asimmetria è prossimo a zero, mentre per i pozzi 14 e 19, come risultato nella stima della distribuzione normale asimmetrica, α tende a ∞ . Analizzando invece il parametro che stima i gradi di libertà per quindici pozzi la distribuzione è approssimabile alla normale asimmetrica.

La Tabella 4.2 riporta i valori ottenuti tramite il metodo di stima *mst.mle*.

Pozzo	ξ	ω	α	ν	$\log L$
1	-0.5904	0.2223	3.698×10^{-1}	3.516	-7.0130
2	-0.7703	0.1508	-2.588×10^{-1}	1.672	-6.3447
3	-0.3576	0.3527	-2.076	6.421×10^5	0.1402
4	-1.3412	0.3281	2.050	2.074	-18.0886
5	-1.2260	0.3350	4.705×10^{-1}	2.598	-25.9579
6	-1.2196	0.3654	3.201	1.312×10^4	2.5411
7	-1.3372	0.5991	4.066	1.133×10^4	-14.6441
8	-0.1821	0.8244	-3.593	4.443×10^1	-20.2216
9	-1.8789	0.4159	4.899×10^{-1}	2.856×10^4	-18.5584
10	-0.8941	0.1245	2.651×10^{-1}	2.917	12.4089
11	-2.0325	0.4131	1.182	5.974×10^3	-11.4280
12	-1.7114	0.3095	-3.098×10^{-1}	3.199	-21.5352
13	-0.9701	0.1385	2.119×10^{-2}	1.182×10^5	21.7500
14	-1.1147	0.5649	3.564×10^2	∞	-5.9279
15	-0.8537	0.1792	-1.136	7.857	14.6717
16	-1.4479	0.3554	2.012	2.691×10^5	-0.3892
17	-2.0235	0.3663	-2.136	9.946	-5.1950
18	-1.7094	0.3165	9.233×10^{-1}	1.655×10^4	-3.5712
19	-0.7062	1.0961	-1.824×10^4	∞	-31.8833
20	-1.2173	0.1184	-9.036×10^{-3}	1.519	-0.2468
21	-1.7982	0.3546	3.563	2.839×10^4	4.2743
22	-0.8960	0.1764	-2.527	1.025×10^4	27.6155
23	-0.9297	0.2695	3.713	1.432	-15.1130
24	-2.2628	0.3842	-1.335	2.218	-26.2200
25	-1.5090	0.5117	1.481	1.285×10^4	-15.8815
26	-0.9676	0.2328	1.001×10^{-2}	4.221×10^3	1.3813

Tabella 4.2 Stime dei parametri ottenute tramite la funzione *mst.mle* e valore della log-verosimiglianza nel massimo.

Confrontando questi risultati con quelli ottenuti in precedenza si traggono le stesse conclusioni nonostante si riscontrino alcune differenze nelle stime del parametro di forma e dei gradi di libertà. Infatti le stime del parametro di asimmetria differiscono per i pozzi 9, 13, 14, 19, 26, ma si nota che tali diversità non risultano significative in quanto si verificano solo per valori di α che tendono a zero (per cui si riottiene la distribuzione *t* di Student) o a ∞ e non portando così ad alterazioni delle conclusioni già fatte. Stesse considerazioni per i gradi di libertà: le differenze si riscontrano solo per $\nu \rightarrow \infty$, ed entrambi i metodi di stima indicano per questi pozzi l'approssimazione della distribuzione alla normale asimmetrica.

Nei casi in cui, invece, le stime dei parametri della distribuzione *t* asimmetrica sono finite le stime ottenute tramite i due metodi coincidono.

Vengono rilevate anche alcune diversità per le stime di ξ e ω (pozzi 9, 13, 14, 19,

26) ma non risultano essere rilevanti. Il valore della log-verosimiglianza differisce solo in due casi, ma in modo non marcato.

A questo punto risulta indifferente la scelta del metodo da utilizzare per la stima dei parametri della distribuzione. Dato che la funzione *st.mle* richiama al suo interno *mst.mle*, nelle analisi che seguono verrà utilizzata quest'ultima.

4.2 Confronto tra distribuzione normale asimmetrica e t asimmetrica

Dalle analisi svolte finora è stato verificato che sia la distribuzione normale asimmetrica che t asimmetrica sono utilizzabili per spiegare la variabile oggetto di studio. In questo paragrafo si vuole capire se è possibile affermare che una è più adeguata dell'altra o è possibile constatare se la distribuzione dei pozzi dipende dalla posizione geografica in cui sono posizionati. Innanzitutto si procede tramite rappresentazioni grafiche che mettono a confronto le curve stimate tramite la funzione *sn.mle* (linea tratteggiata) e *mst.mle* (linea continua).

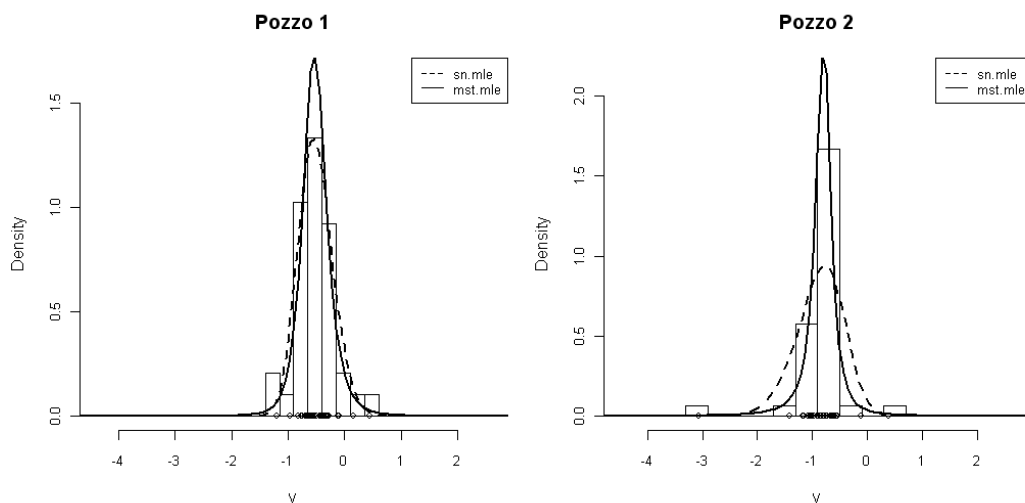


Figura 4.2 Pozzi 1-2 Istogramma con le curve della densità normale asimmetrica stimata tramite *sn.mle* (linea tratteggiata) e della densità t asimmetrica tramite *mst.mle* (linea continua).

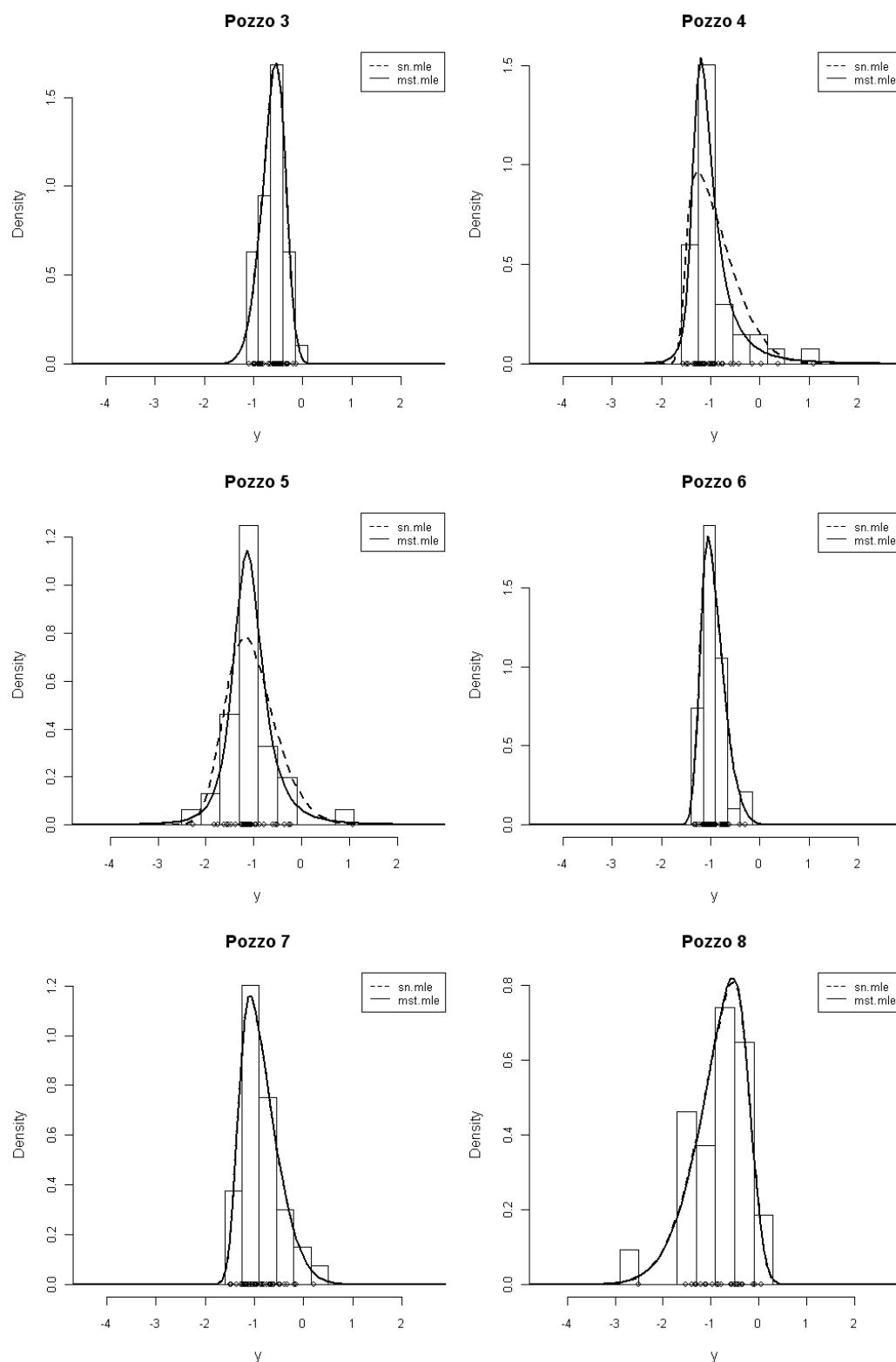


Figura 4.2 Pozzi 3-8 Istogramma con le curve della densità normale asimmetrica stimata tramite *sn.mle* (linea tratteggiata) e della densità *t* asimmetrica tramite *mst.mle* (linea continua).

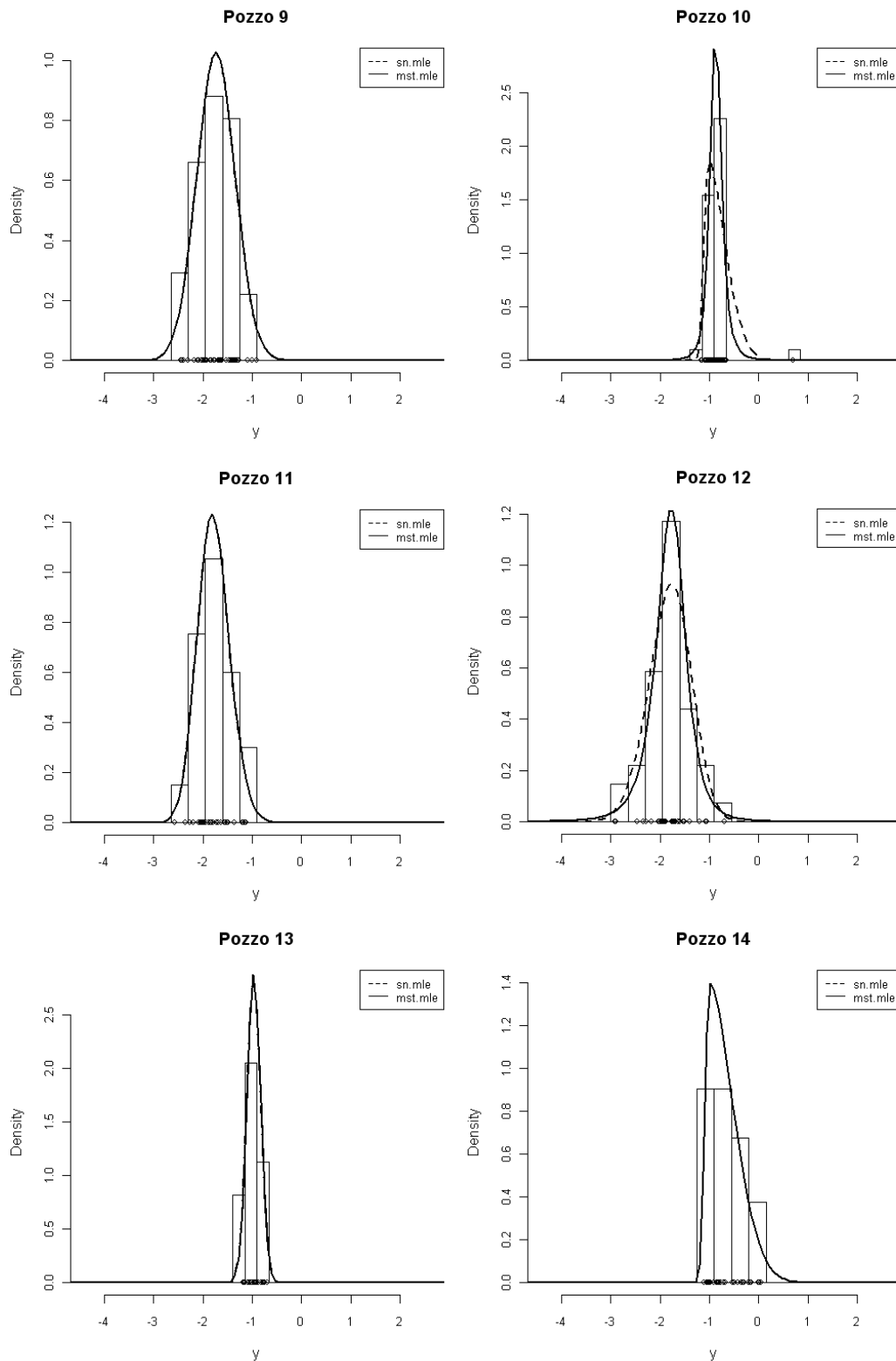


Figura 4.2 Pozzi 9-14 Istogramma con le curve della densità normale asimmetrica stimata tramite $sn.mle$ (linea tratteggiata) e della densità t asimmetrica tramite $mst.mle$ (linea continua).

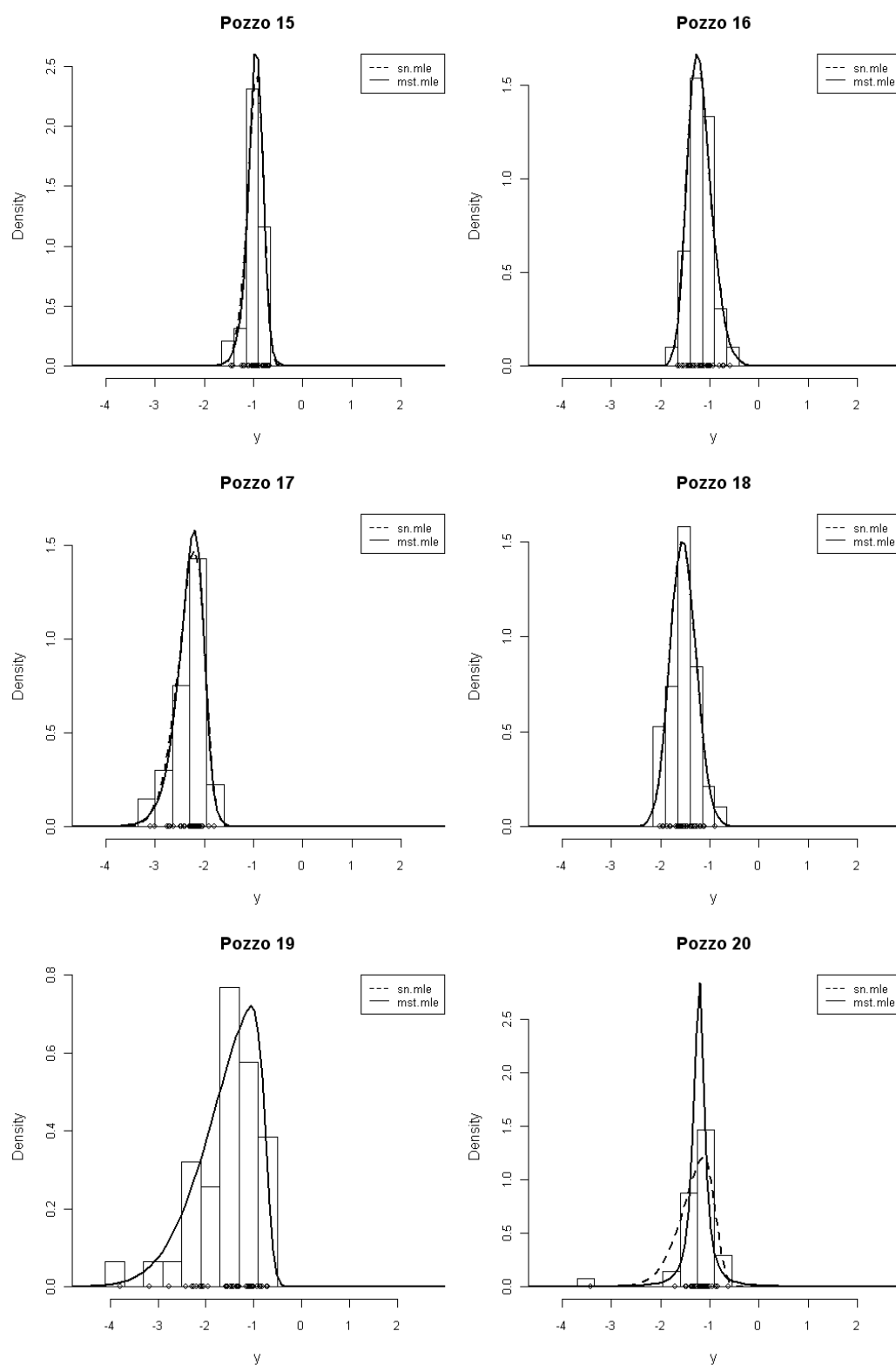


Figura 4.2 Pozzi 15-20 Istogramma con le curve della densità normale asimmetrica stimata tramite *sn.mle* (linea tratteggiata) e della densità *t* asimmetrica tramite *mst.mle* (linea continua).

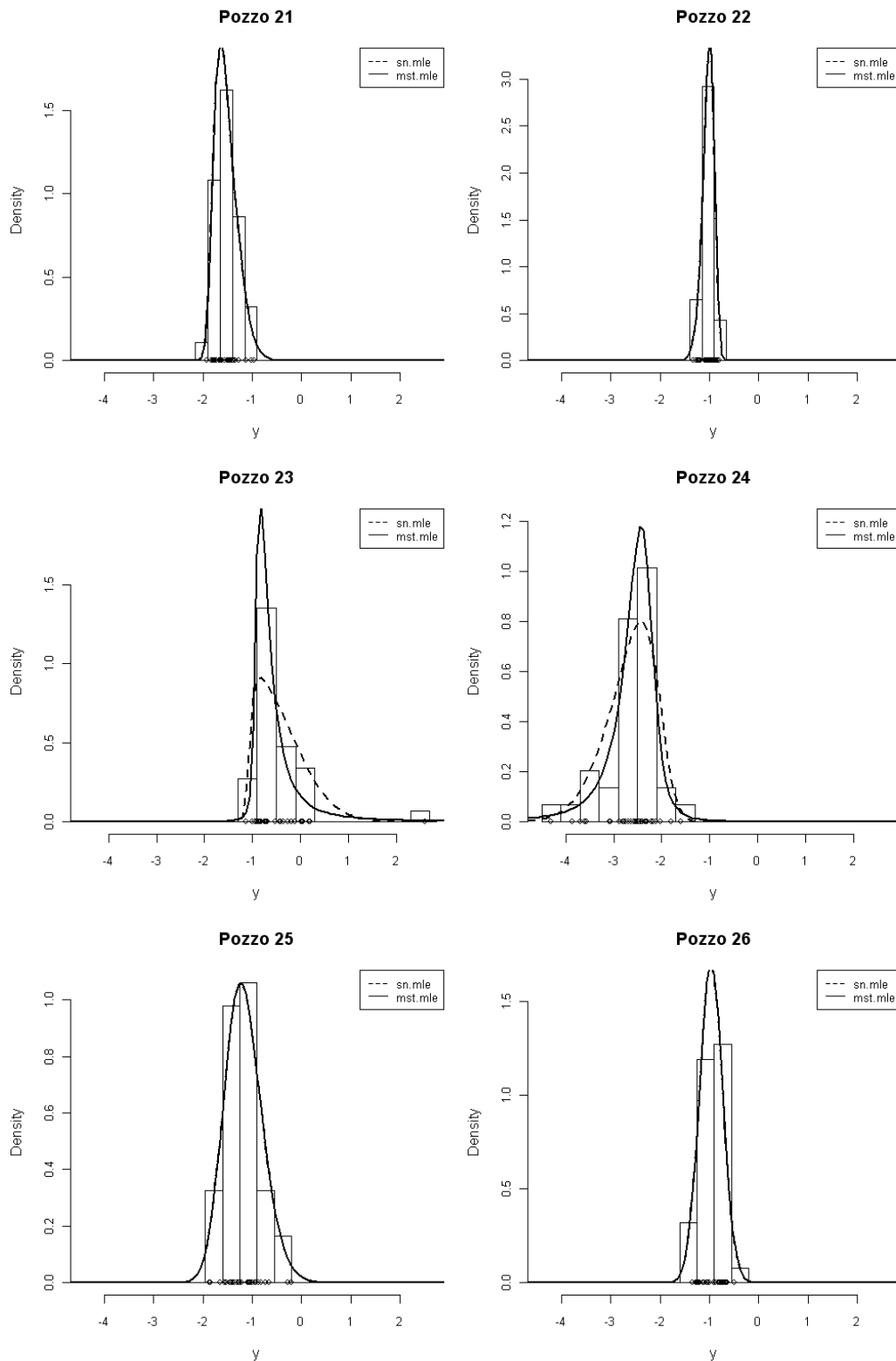


Figura 4.2 Pozzi 21-26 Istogramma con le curve della densità normale asimmetrica stimata tramite *sn.mle* (linea tratteggiata) e della densità t asimmetrica tramite *mst.mle* (linea continua).

I grafici riportati permettono di avere conferma delle considerazioni fatte in precedenza e di individuare i pozzi la cui distribuzione è approssimabile alla normale asimmetrica. Infatti, nel caso in cui la stima di $\nu \rightarrow \infty$, le due curve disegnate coincidono e ciò avviene per quindici pozzi individuabili anche attraverso la Tabella 4.2 (3, 6, 7, 8, 9, 11, 13, 14, 16, 18, 19, 21, 22, 25, 26).

Negli altri casi la curva della distribuzione t asimmetrica segue meglio l'andamento dell'istogramma nel punto in cui la densità della classe risulta superiore.

Per verificare se è possibile identificare un'associazione tra la posizione dei pozzi e la distribuzione della variabile, si considera il parametro che descrive i gradi di libertà della t e si procede tramite il meccanismo della regressione locale utilizzato anche per analizzare il parametro di asimmetria della distribuzione normale asimmetrica.

4.3 Stima del parametro dei gradi di libertà tramite regressione locale

Per capire se la posizione geografica influisce sul valore della stima dei gradi di libertà, e quindi sulla distribuzione della variabile, si applica il meccanismo della regressione locale su tale parametro.

Si deve tenere conto, però, che la stima assume valori in un intervallo molto ampio, $1.432 \leq \nu < \infty$, ed è necessario riportarlo in una scala che consenta di apprezzare maggiormente le varie sfumature delle differenti stime. Il modo più appropriato è considerare il logaritmo di questo parametro.

Nella Tabella 4.3 vengono riportate le corrispondenze tra la stima e la relativa trasformazione logaritmica.

Per effettuare le rappresentazioni grafiche si deve tenere conto di una problematica: nonostante l'applicazione del logaritmo per due pozzi $\nu = \infty$; per ovviare a

questa situazione, si sostituisce la stima ottenuta con il valore massimo di questo parametro preso tra gli altri pozzi.

Pozzo	ν	$\log(\nu)$
1	3.52	1.257
2	1.67	0.514
3	6.42×10^5	13.372
4	2.07	0.729
5	2.60	0.955
6	1.31×10^4	9.482
7	1.13×10^4	9.335
8	4.44×10^1	3.794
9	2.86×10^4	10.260
10	2.92	1.071
11	5.97×10^3	8.695
12	3.20	1.163
13	1.18×10^5	11.680

Pozzo	ν	$\log(\nu)$
14	Inf	13.372
15	7.86	2.061
16	2.69×10^5	12.503
17	9.95	2.297
18	1.65×10^4	9.714
19	Inf	13.372
20	1.52	0.418
21	2.84×10^4	10.254
22	1.02×10^4	9.235
23	1.43	0.359
24	2.22	0.797
25	1.29×10^4	9.461
26	4.22×10^3	8.348

Tabella 4.3 Corrispondenza tra ν e $\log(\nu)$.

Il grafico rappresentato permette di distinguere alcune zone in base al valore assunto dal parametro considerato.

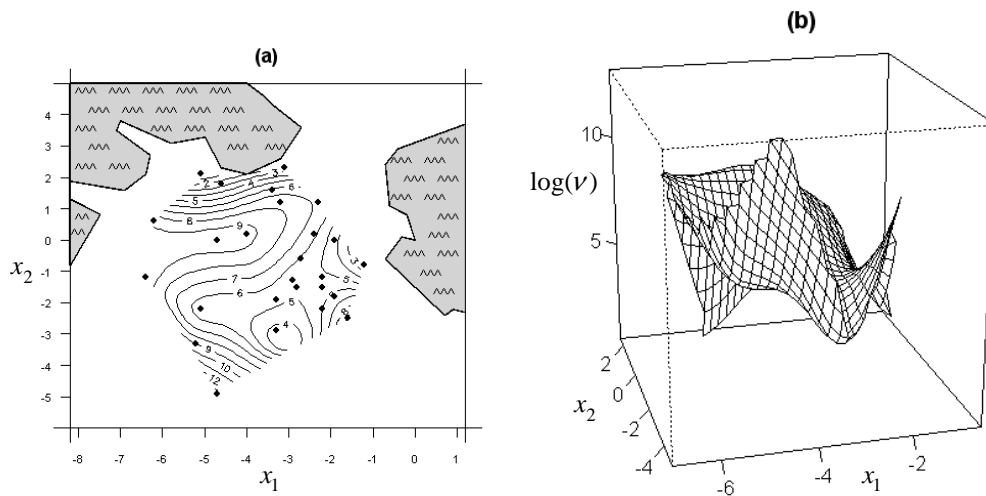


Figura 4.3 Curve di livello (a) e grafico prospettico (b) del logaritmo della stima dei gradi di libertà tramite la regressione locale con due variabili esplicative espresse dalle coordinate dei punti che localizzano i pozzi.

Innanzitutto è possibile notare che i valori più bassi sono collocati nei territori prossimi alle coste del Porto di Ponente e Porto di Levante; quindi, per i pozzi localizzati in queste zone, la variabile studiata può essere analizzata tramite la di-

sistribuzione t asimmetrica. Man mano ci si avvicina al vulcano i gradi di libertà aumentano e portano all'utilizzo della normale asimmetrica per esaminare la variabile.

Tale situazione è confermata anche dal grafico prospettico in Figura 4.3 (b) in cui è visibile in maniera evidente che i punti più bassi sono localizzati nei valori più alti dell'asse delle ordinate (x_2) indicanti le coste dei due Porti. Spostandosi verso il vulcano, dopo una leggera crescita dei valori si riscontra una diminuzione maggiormente accentuata nella parte centrale della zona inferiore; al contrario, le fasce esterne inferiori rappresentano i picchi più alti

Si esaminano ora le modifiche che subiscono le curve di livello al variare del parametro h , che stabilisce il livello di lisciamento della curva stimata. Per il grafico in Figura 4.3 (a) tale valore è stato fissato dal programma pari a $h_1 = h_2 = 0.898$; le rappresentazioni in Figura 4.4 riportano le differenze che si registrano diminuendo ($h=0.5$) e aumentando ($h=1.2$) il valore assegnato al parametro di lisciamento.

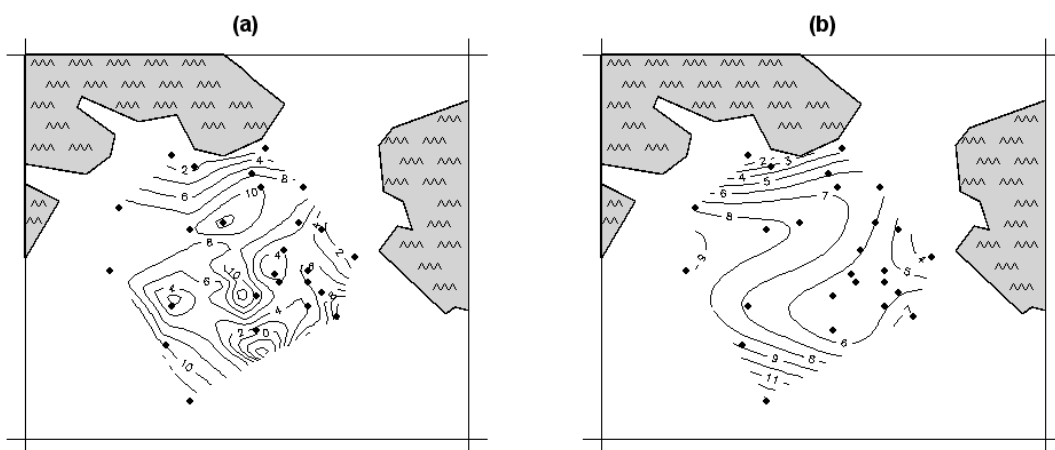


Figura 4.4 Curve di livello della stima della mediana tramite la regressione locale con due variabili esplicative espresse dalle coordinate dei punti che localizzano i pozzi, fissando $h=0.5$ (a) e $h=1.2$ (b).

Nei grafici sopra riportati si riconosce lo stesso andamento di fondo riscontrato in Figura 4.3 nonostante in (b) le curve presentino minori fluttuazioni dovute

all'aumento del valore di h e in (a) la stima risulti più accurata (h più basso). In quest'ultimo pannello, a differenza degli altri, le curve sono più aderenti al comportamento dei dati ed è possibile individuare più punti di massimo nella parte centrale della zona considerata, evidenziando con maggiore precisioni le oscillazioni riscontrate nella Figura 4.3 (b).

Si conclude che il parametro che indica i gradi di libertà della t asimmetrica risulta un fattore discriminante per individuare la distribuzione che segue meglio l'andamento dei dati osservati per ogni pozzo; infatti, note le coordinate che lo localizza, si può conoscere se il parametro assume valore sufficientemente alto per poter approssimare la distribuzione alla normale asimmetrica.

4.4 Stima della mediana tramite regressione locale

Come già eseguito durante l'analisi della distribuzione normale asimmetrica, si vuole capire se esiste un'associazione tra la localizzazione dei pozzi e la relativa mediana, indice di posizione utilizzato per sintetizzare la variabile analizzata in un unico valore per ogni pozzo, utilizzando il metodo della regressione locale.

Pozzo	Mediana	Pozzo	Mediana
1	-0.525	14	-0.734
2	-0.805	15	-0.960
3	-0.590	16	-1.215
4	-1.091	17	-2.273
5	-1.102	18	-1.543
6	-0.974	19	-1.445
7	-0.933	20	-1.218
8	-0.742	21	-1.559
9	-1.734	22	-1.014
10	-0.867	23	-0.697
11	-1.792	24	-2.526
12	-1.790	25	-1.190
13	-0.968	26	-0.966

Tabella 4.4 Mediana delle osservazioni registrate per ogni pozzo.

Confrontando i risultati sopra riportati con quelli riportati in Tabella 3.13, per alcuni pozzi i valori coincidono; tale situazione si verifica in corrispondenza dei

quali la stima di ν è risultata sufficientemente alta per rendere valida l'approssimazione della distribuzione t asimmetrica alla normale asimmetrica

Visto il rilevante numero di pozzi la cui stima del parametro dei gradi di libertà tende a ∞ , non stupisce la notevole somiglianza dei grafici sotto riportati con quelli ottenuti nell'analisi svolta nel paragrafo 3.5.

In Figura 4.5 (a), partendo dalla parte inferiore del grafico, si riscontrano stime che in valore assoluto decrescono fino alla fascia centrale, punto da cui crescono nuovamente man mano ci si avvicina alla costa del Porto di Ponente. Questa situazione è confermata dal grafico in (b) in cui le stime risultanti presentano forma parabolica con concavità verso il basso e il vertice è in corrispondenza alla parte centrale della zona esaminata.

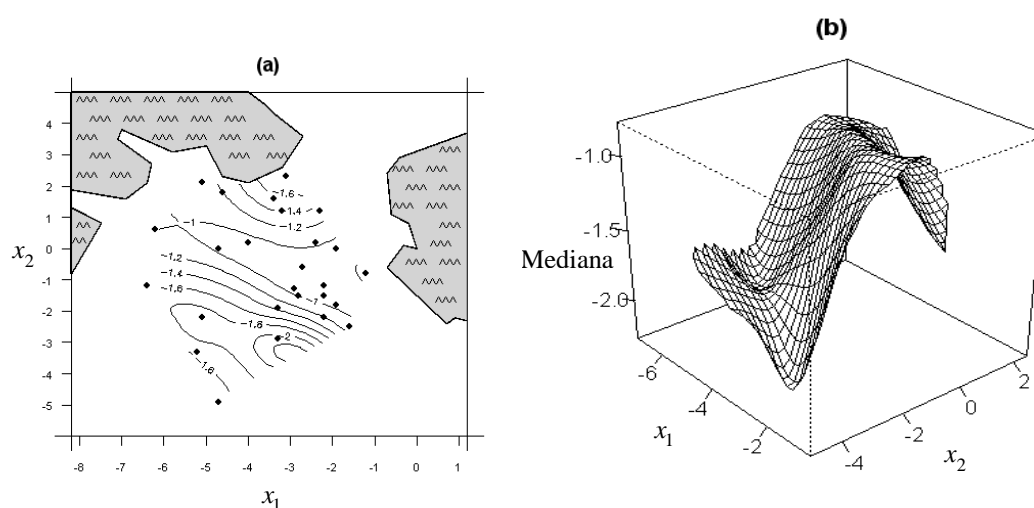


Figura 4.5 Curve di livello (a) e grafico prospettico (b) della stima della mediana tramite la regressione locale con due variabili esplicative espresse dalle coordinate dei punti che localizzano i pozzi.

Di seguito si verificano le modifiche che subiscono le curve di livello al variare del parametro h che stabilisce il livello di liscio della curva stimata. Come nel caso appena esaminato ci si aspetta risultino rappresentazioni grafiche molto simili a quelle già analizzate per la normale asimmetrica

Per i grafici sopra riportati il valore di questo parametro è stato fissato dal programma pari a $h_1 = h_2 = 0.898$; le rappresentazioni in Figura 4.6 riportano le dif-

Per meglio valutare i grafici, le stime risultate negative, verranno trattate in valore

Nel pannello di sinistra (valore di h più piccolo) si riconosce lo stesso comporta-

Le stesse considerazioni vengono fatte anche per il pannello di destra, anche se i

(Si ricorda che queste conclusioni riferiscono alle stime considerate in valore as-

Dai risultati ottenuti si può affermare che come per la normale asimmetrica

importante per capire per ognuno di essi la distribuzione che segue meglio l'andamento della variabile e i valori assunti da questa nel descrivere il rapporto tra catione calcio e anione solfato.

4.5 Stima della mediana tramite verosimiglianza locale

Come affrontato nel paragrafo 3.6, analizzando la stima della mediana tramite il metodo della regressione locale si ignora la natura della variabile considerata. Per rimediare a questa problematica è necessario applicare il meccanismo della verosimiglianza locale ottenuta tramite l'applicazione di pesi alla log-verosimiglianza nella forma

$$l_{[h_1, h_2, x_{01}, x_{02}]}(\vartheta) = \sum_{i=1}^m l_i(\vartheta) w(x_{i1} - x_{01}; h_1) w(x_{i2} - x_{02}; h_2)$$

dove $\vartheta = (\xi, \omega, \alpha, \nu)$. Massimizzando la log-verosimiglianza rispetto a ϑ si ottengono le stime dei parametri della distribuzione t asimmetrica del pozzo di coordinate x_0 .

Dopo aver creato una funzione per il calcolo della verosimiglianza pesata, si confrontano i valori dei parametri stimati per ogni pozzo con questo procedimento e tramite la funzione *mst.mle* utilizzata nel paragrafo 4.1.

Come per la normale asimmetrica, dalle tabelle riportate è possibile verificare che, nonostante l'effetto del parametro di lisciamento sia stato in gran parte annullato ponendo $h=0.1$, l'aggiunta di pesi alla log-verosimiglianza produce differenze nelle stime dei parametri; in particolare le variazioni si registrano per pozzi che non si presentano in posizione isolata rispetto agli altri (per la localizzazione si fa riferimento alla Figura 1.2).

<i>Pozzo</i>	ξ	ω	α	ν
1	-0.5904	0.2223	3.698×10^{-1}	3.516
2	-0.7703	0.1508	-2.588×10^{-1}	1.672
3	-0.3576	0.3527	-2.076	6.421×10^5
4	-1.3412	0.3281	2.050	2.074
5	-1.2260	0.3350	4.705×10^{-1}	2.598
6	-1.2196	0.3654	3.201	1.312×10^4
7	-1.3372	0.5991	4.066	1.133×10^4
8	-0.1821	0.8244	-3.593	4.443×10^1
9	-1.8789	0.4159	4.899×10^{-1}	2.856×10^4
10	-0.8941	0.1245	2.651×10^{-1}	2.917
11	-2.0325	0.4131	1.182	5.974×10^3
12	-1.7114	0.3095	-3.098×10^{-1}	3.199
13	-0.9701	0.1385	2.119×10^{-2}	1.182×10^5
14	-1.1147	0.5649	3.564×10^2	∞
15	-0.8537	0.1792	-1.136	7.857
16	-1.4479	0.3554	2.012	2.691×10^5
17	-2.0235	0.3663	-2.136	9.946
18	-1.7094	0.3165	9.233×10^{-1}	1.655×10^4
19	-0.7062	1.0961	-1.824×10^4	∞
20	-1.2173	0.1184	-9.036×10^{-3}	1.519
21	-1.7982	0.3546	3.563	2.839×10^4
22	-0.8960	0.1764	-2.527	1.025×10^4
23	-0.9297	0.2695	3.713	1.432
24	-2.2628	0.3842	-1.335	2.218
25	-1.5090	0.5117	1.481	1.285×10^4
26	-0.9676	0.2328	1.001×10^{-2}	4.221×10^3

<i>Pozzo</i>	ξ	ω	α	ν
1	-0.5943	0.2271	2.649×10^{-1}	3.413
2	-0.7778	0.1598	-8.990×10^{-2}	1.720
3	-0.6238	0.2437	5.669×10^{-2}	6.081×10^5
4	-1.3412	0.3281	2.050	2.074
5	-1.2260	0.3350	4.705×10^{-1}	2.598
6	-1.2196	0.3654	3.201	1.049×10^5
7	-1.3372	0.5993	4.067	8.624×10^5
8	-0.1821	0.8244	-3.593	4.443×10^1
9	-1.7242	0.3895	-2.789×10^{-2}	7.220×10^5
10	-0.8941	0.1245	2.651×10^{-1}	2.917
11	-2.0217	0.4065	1.111	3.070×10^6
12	-1.7114	0.3095	-3.098×10^{-1}	3.199
13	-0.9788	0.1390	9.957×10^{-2}	7.002×10^5
14	-1.1113	0.5435	1.504×10^1	2.587×10^1
15	-0.9120	0.1390	-4.637×10^{-1}	4.280
16	-1.4499	0.3564	2.055	7.847×10^5
17	-2.0235	0.3663	-2.136	9.946
18	-1.6842	0.3042	7.535×10^{-1}	4.058×10^6
19	-0.2479	1.1567	-1.332×10^2	2.003×10^2
20	-1.2173	0.1184	-9.033×10^{-3}	1.519
21	-1.7982	0.3546	3.562	4.733×10^4
22	-0.8961	0.1763	-2.520	2.194×10^3
23	-0.9297	0.2695	3.713	1.432
24	-2.2628	0.3842	-1.335	2.218
25	-1.5090	0.5118	1.481	1.659×10^5
26	-1.0071	0.2365	2.245×10^{-1}	4.770×10^4

(a) Parametri stimati tramite verosimiglianza locale ponendo $h=0.1$.

(b) Parametri stimati tramite la funzione *mst.mle*.

Tabella 4.5 Riassunto dei valori ottenuti tramite i due metodi di stima considerati.

Visto che l'aggiunta di pesi produce differenze nella log-verosimiglianza si procede analizzando le diversità che si riscontrano nella stima della mediana tramite la verosimiglianza locale e la regressione locale variando il valore del parametro di lisciamiento. Come spiegato nell'applicazione di questo meccanismo

durante l'analisi della normale asimmetrica (pag.53), la stima non viene limitata ai punti osservati ma la si estende ad una griglia posizionata in un intervallo che ricopre i ventisei punti osservati; successivamente, tramite la verosimiglianza locale, sono stati stimati i parametri della distribuzione per ogni punto di intersezione nella griglia (400 punti). Infine, per la rappresentazione grafica, sono state disegnate le curve di livello della mediana relativa alle stime calcolate secondo diversi livelli di liscio.

Si procede, quindi, analizzando le differenze che si riscontrano nella stima della mediana tramite la verosimiglianza locale e la regressione locale variando il valore del parametro di liscio.

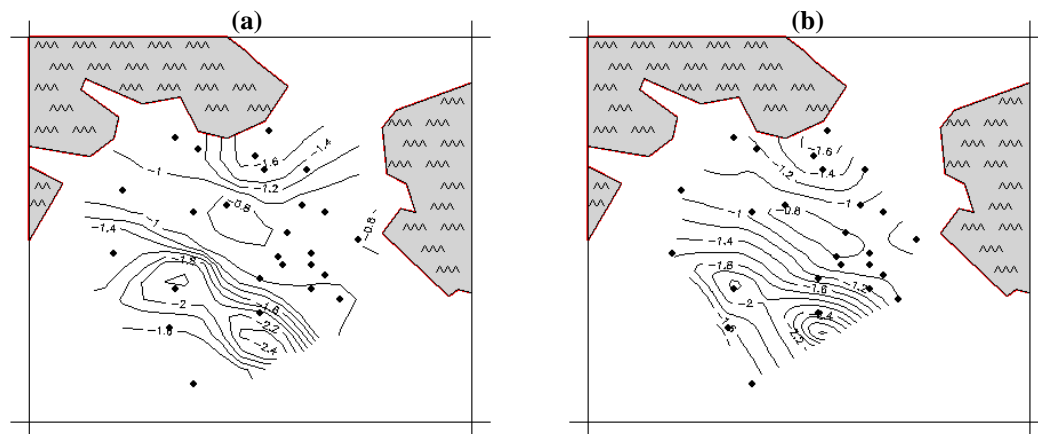


Figura 4.7 Curve di livello della stima della mediana tramite la verosimiglianza locale (a) e regressione locale (b) con due variabili esplicative espresse dalle coordinate dei punti che localizzano i pozzi, fissando $h=0.6$.

Questi grafici presentano notevoli somiglianze con quelli riportati in Figura 3.16 (a-b). Come riscontrato nel paragrafo precedente, l'elevata numerosità di pozzi in cui $\nu \rightarrow \infty$ spiega le similarità che si ottengono analizzando la distribuzione t asimmetrica rispetto alla normale asimmetrica.

Per verificarlo si esaminano i principali indici di posizione che riassumono i valori assunti dalle stime dei gradi di libertà.

Minimo	1°Quartile	Mediana	Media	3°Quartile	Massimo
1.325	2.712	5.896	7.997×10^6	2.055×10^1	2.079×10^9

Tabella 4.6 Caratteristiche dei valori assunti dalle stime dei gradi di libertà ponendo $h=0.6$.

L'intervallo in cui si estendono le stime è molto ampio e ciò spiega perché media e mediana assumono valori così diversi. Il valore basso della media indica la presenza di stime di gradi di libertà che tendono a infinito ma il valore della mediana fa presupporre che le stime così alte non siano molto numerose.

Se si pone $\nu=30$ come soglia per individuare quando è possibile approssimare la distribuzione t asimmetrica alla normale asimmetrica, 91 stime su 400 risultano superarla e quindi rendere valida tale approssimazione.

Le medesime conclusioni valgono anche per $h=1$ come dimostrano i grafici e la tabella sottostanti.

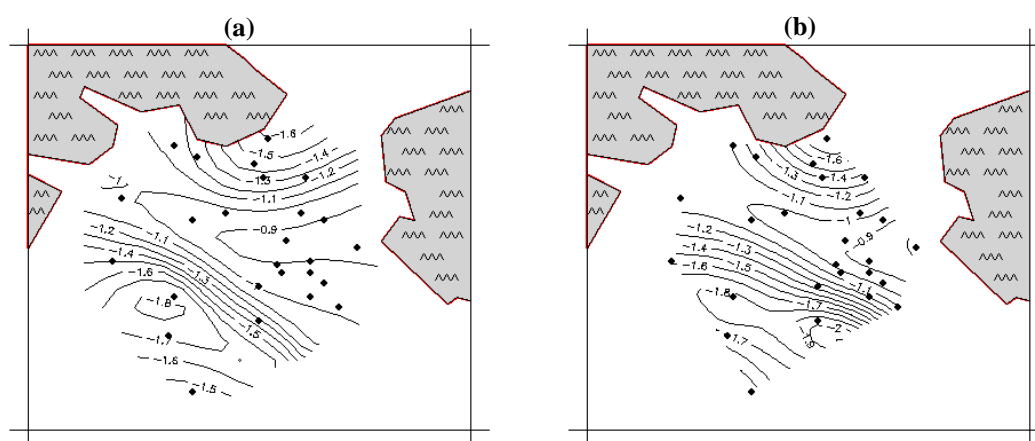


Figura 4.8 Curve di livello della stima della mediana tramite la verosimiglianza locale (a) e regressione locale (b) con due variabili esplicative espresse dalle coordinate dei punti che localizzano i pozzi, fissando $h=1$.

Minimo	1°Quartile	Mediana	Media	3°Quartile	Massimo
1.322	2.256	5.604	7.888×10^6	1.822×10^1	3.134×10^9

Tabella 4.7 Caratteristiche dei valori assunti dalle stime dei gradi di libertà ponendo $h=1$.

Aumentano il valore del parametro di lisciamento le curve di livello seguono in maniera meno dettagliata i dati osservati rispetto al caso precedente. Le rappresentazioni grafiche però assomigliano molto a quelle in Figura 3.16 (c-d) e la similarità alla normale asimmetrica è confermata dalle caratteristiche dei valori assunti dalle stime dei gradi di libertà.

Aumentando il valore del parametro di liscio si traggono le medesime conclusioni sia per le rappresentazioni grafiche che per le stime dei gradi di libertà.

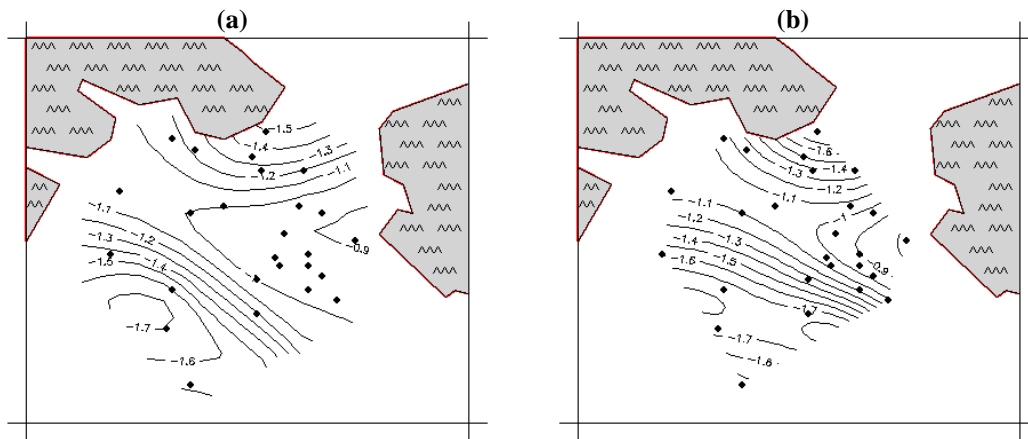


Figura 4.9 Curve di livello della stima della mediana tramite la verosimiglianza locale (a) e regressione locale (b) con due variabili esplicative espresse dalle coordinate dei punti che localizzano i pozzi, fissando $h=1.2$.

Minimo	1°Quartile	Mediana	Media	3°Quartile	Massimo
1.323	2.201e	5.670	1.681×10^4	1.473×10^1	1.529×10^6

Tabella 4.8 Caratteristiche dei valori assunti dalle stime dei gradi di libertà ponendo $h=1.2$.

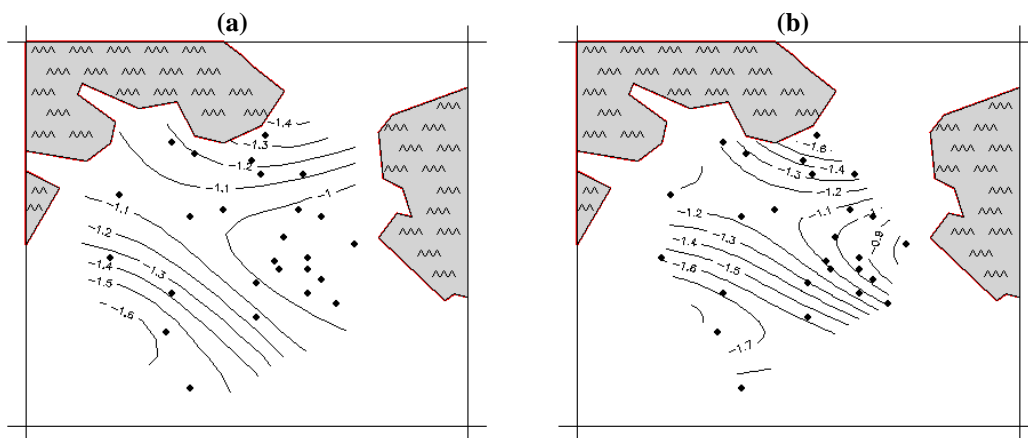


Figura 4.10 Curve di livello della stima della mediana tramite la verosimiglianza locale (a) e regressione locale (b) con due variabili esplicative espresse dalle coordinate dei punti che localizzano i pozzi, fissando $h=1.5$.

Minimo	1°Quartile	Mediana	Media	3°Quartile	Massimo
1.338	2.218	5.296	8.102	11.740	52.850

Tabella 4.9 Caratteristiche dei valori assunti dalle stime dei gradi di libertà ponendo $h=1.5$.

Nonostante all'aumentare del parametro di liscio le stime dei gradi di libertà assumano valori compresi in intervalli sempre più piccoli, rimane alta la somiglianza con i grafici ottenuti durante la stima della mediana tramite verosimiglianza locale nell'analisi della distribuzione normale asimmetrica.

In tutti i grafici è possibile suddividere la zona considerata tre in parti a seconda dell'andamento delle curve. Quella in cui si riscontrano maggiori differenze rispetto alla stima tramite regressione locale è quella inferiore, nella quale i punti di massimo vengono individuati in maniera differente.

Capitolo 5

ANALISI DELLA VARIABILE IN PERIODI DI ATTIVITÀ ANOMALA DEL VULCANO

Dopo l'ultima eruzione nell'isola, verificatasi dal 1888 al 1890, si sono riscontrati vari periodi di attività anomala del vulcano. Incrementi termici, attività sismiche e geochimiche furono registrati nel 1978; nel 1980 le attività fumaroliche del cratere subirono importanti cambiamenti sia nell'estensione dell'area che nel numero di fumarole. Un apparente episodio di agitazione si verificò nel 1985 e 1987-1988, aumentando il rischio di ripresa dell'attività eruttiva.

Dato che è nota la data in cui è stato effettuato il campionamento, espressa in mese e anno nell'arco temporale tra giugno 1977 e novembre 1994, è possibile verificare se vi è una forma di relazione tra la variabile oggetto di studio e i periodi di attività anomala del vulcano.

5.1 Analisi grafica

Una prima analisi prevede la rappresentazione grafica dei valori del log-rapporto tra calcio e solfato di ogni pozzo sotto forma di serie temporale in cui vengono evidenziati i periodi sopra citati. In questo modo si vuole esaminare se è possibile riscontrare un andamento sistematico per tutti i pozzi nelle date indicate.

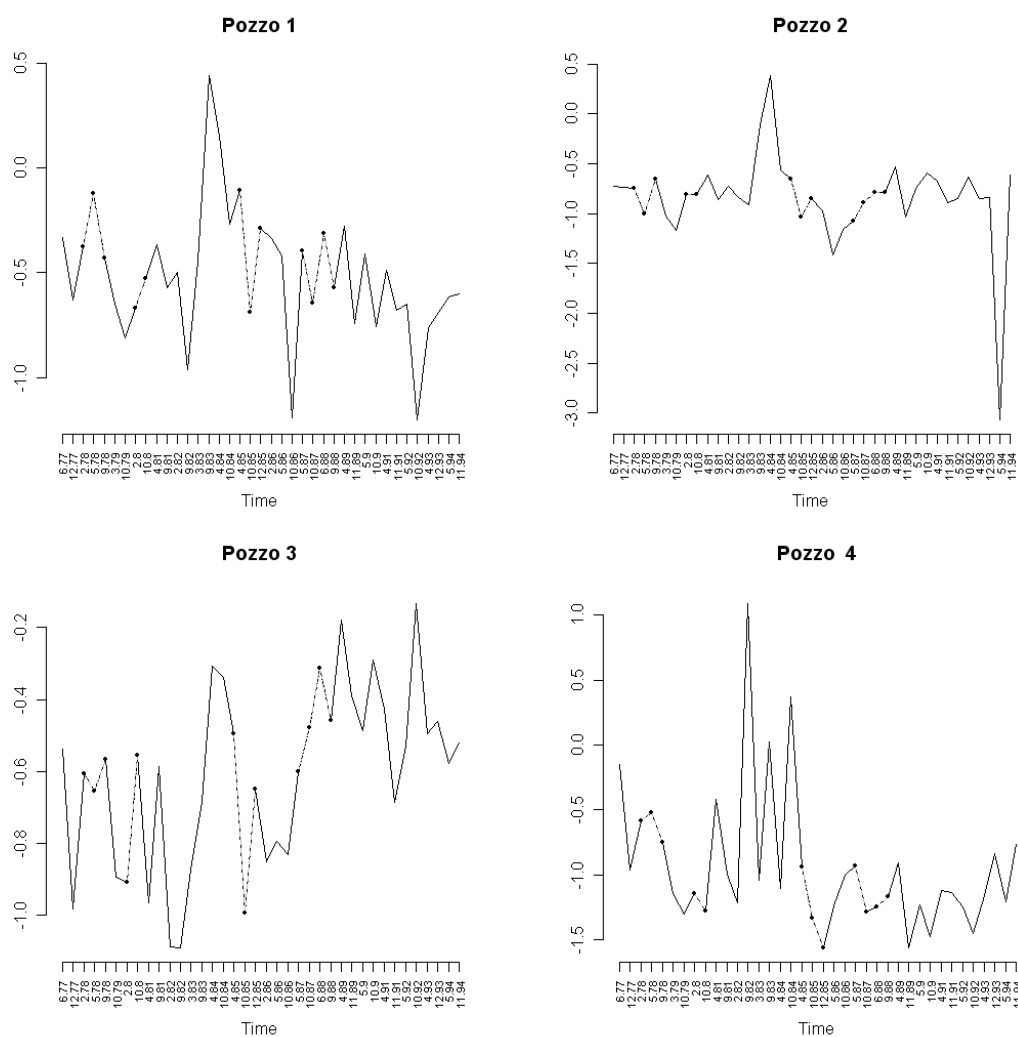


Figura 5.1 Pozzi 1-4 Serie temporale della variabile analizzata mettendo in evidenza i periodi di attività anomala del vulcano.

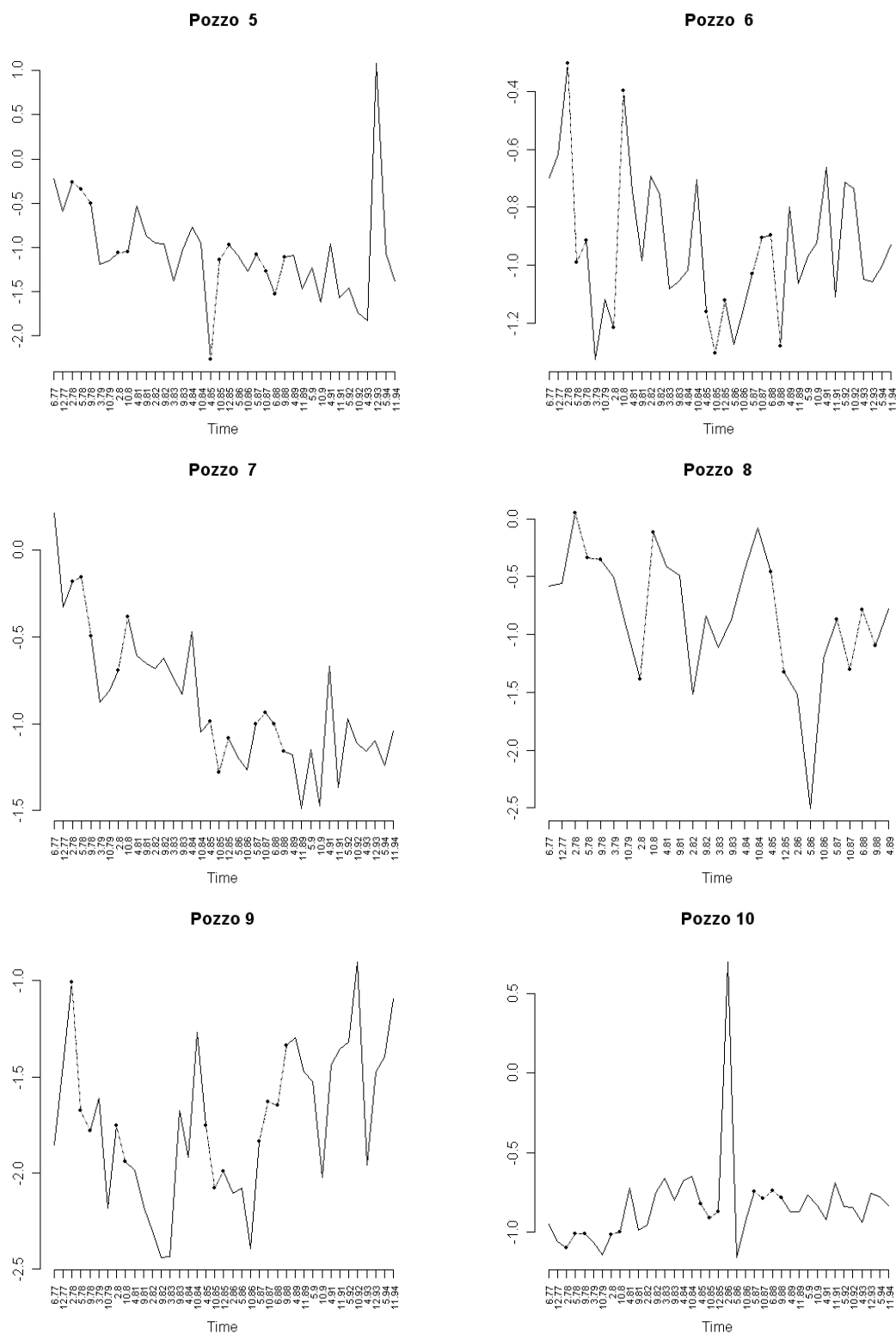


Figura 5.1 Pozzi 5-10 Serie temporale della variabile analizzata mettendo in evidenza i periodi di attività anomala del vulcano.

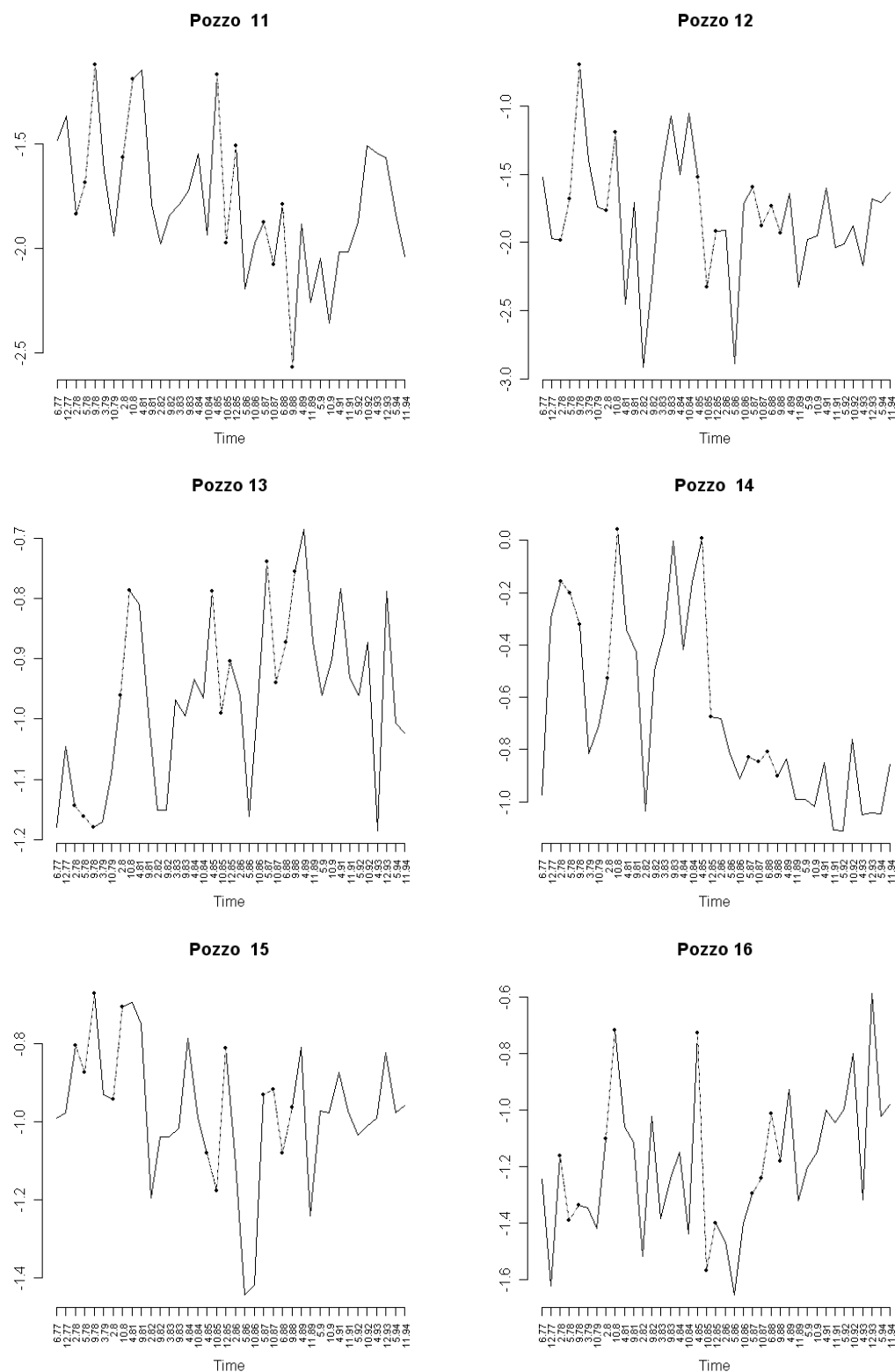


Figura 5.1 Pozzi 11-16 Serie temporale della variabile analizzata mettendo in evidenza i periodi di attività anomala del vulcano.

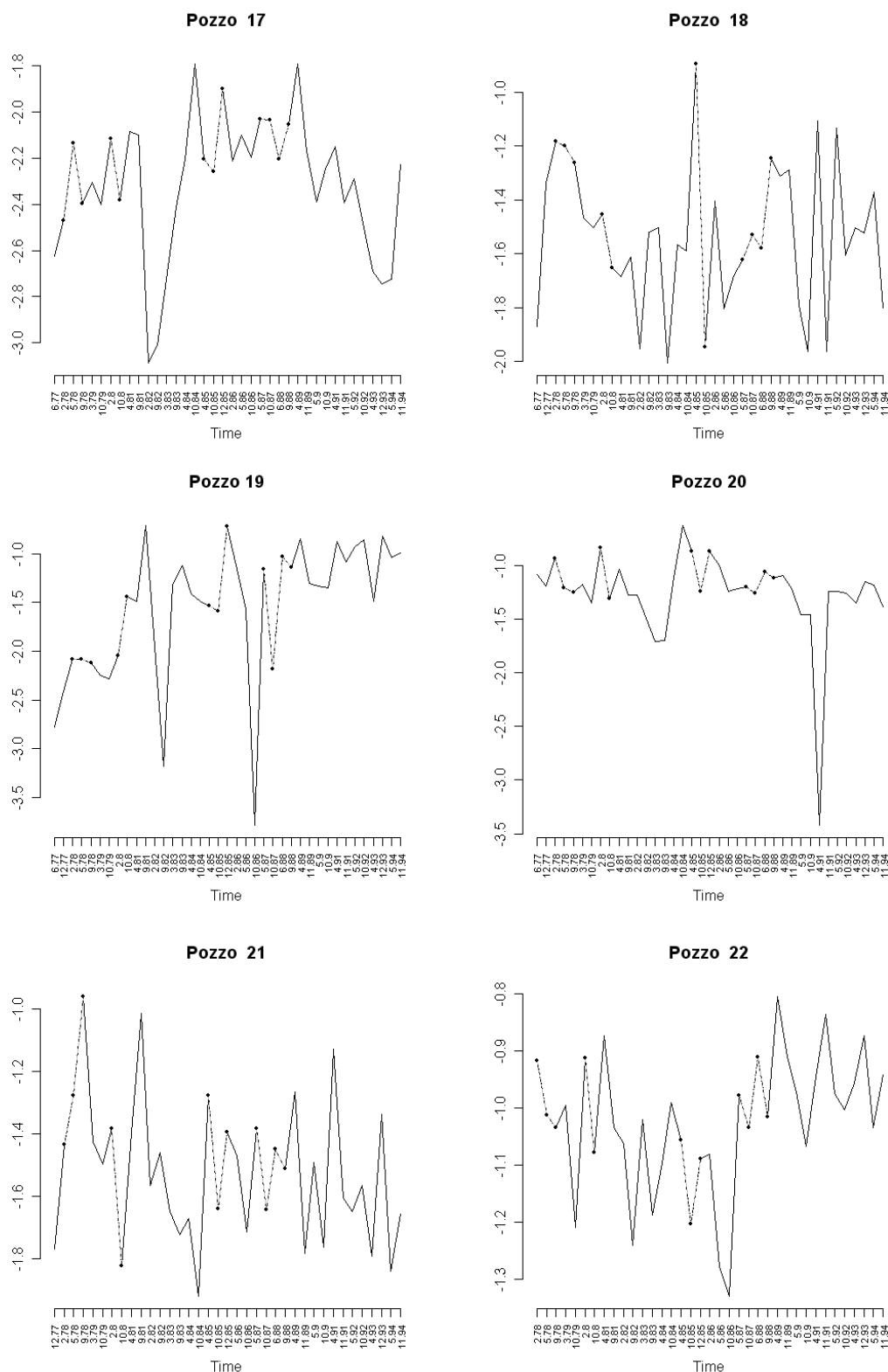


Figura 5.1 Pozzi 17-22 Serie temporale della variabile analizzata mettendo in evidenza i periodi di attività anomala del vulcano.

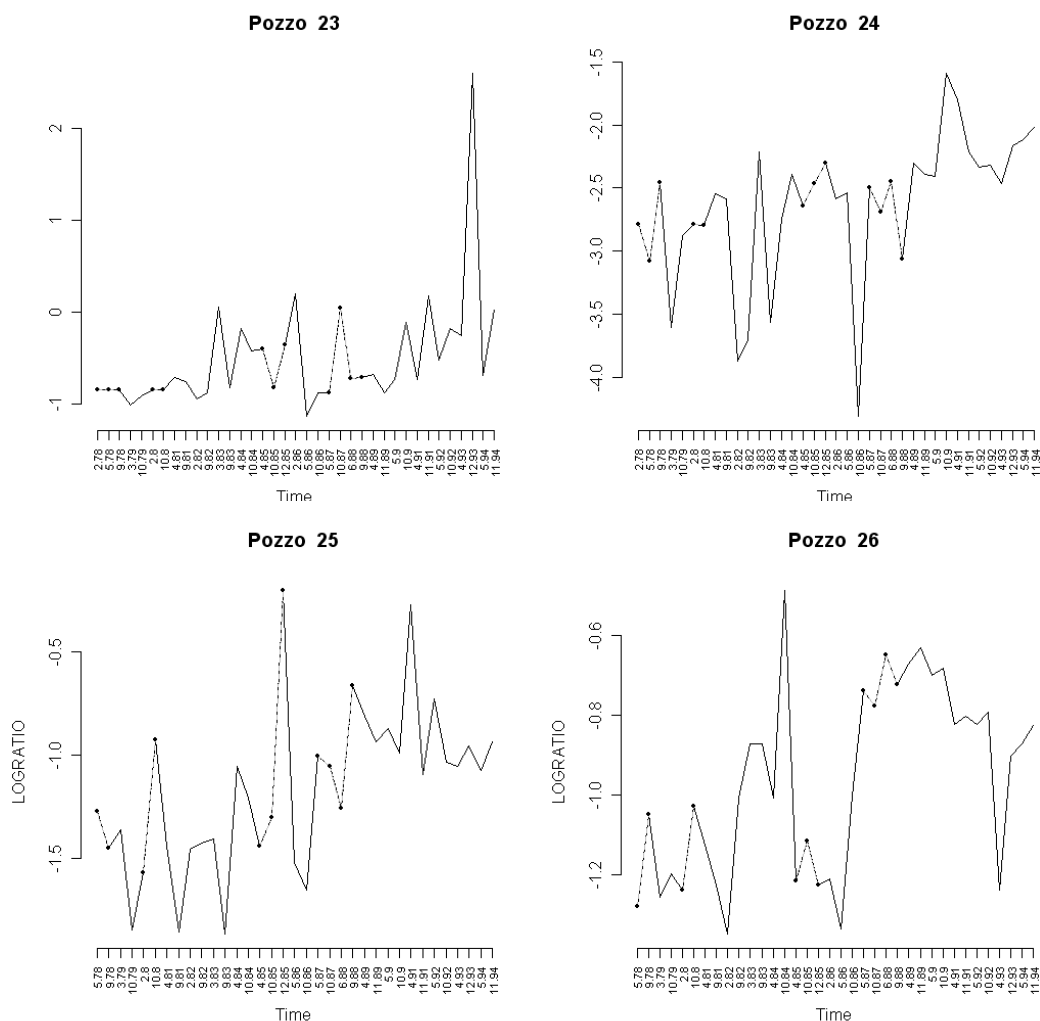


Figura 5.1 Pozzi 23-26 Serie temporale della variabile analizzata mettendo in evidenza i periodi di attività anomala del vulcano.

Nei grafici riportati vengono evidenziati con una linea tratteggiata gli anni in cui sono state registrate attività anomala del vulcano. Analizzando ogni pozzo in tali periodi, però, non è possibile identificare un andamento sistematico univoco dei valori assunti dalla variabile in ognuno di essi, ma per alcuni i valori della variabile incrementano e per altri diminuiscono.

Visto che esaminare il rapporto tra calcio e solfato negli anni indicati non ha portato a utili conclusioni, lo si analizza nei periodi che precedono e seguono gli anni evidenziati. Anche in questo caso, però, nei valori assunti dalla variabile non è possibile identificare un andamento specifico.

L'analisi grafica svolta porta a concludere che il comportamento del rapporto studiato non sembra essere influenzato dai periodi di attività anomala del vulcano.

5.2 Stima della mediana tramite verosimiglianza locale

Questo metodo di stima è già stato utilizzato nei capitoli precedenti, sia durante l'analisi della distribuzione normale asimmetrica che della t asimmetrica. Si vuole ora estendere tale meccanismo per verificare se la data di campionamento (t_1) influisce sui valori assunti dalla variabile: oltre alle coordinate che localizzano il pozzo la log-vero-simiglianza viene moltiplicata anche per un peso calcolato in base alla distanza temporale da una data fissata (t_0) nella forma

$$l_{[h_1, h_2, x_{01}, x_{02}]}(\mathcal{V}) = \sum_{i=1}^m \left[\sum_{j=1}^{n_j} l_{ij}(\mathcal{V}) \right] w(x_{i1} - x_{01}; h_1) w(x_{i2} - x_{02}; h_2) w(t_1 - t_0; h_3)$$

A tale proposito si mettono a confronto i cambiamenti che subisce la stima della mediana in una data in cui è stato registrato un rischio di ripresa di attività eruttiva, ottobre 1985 ($t_0 = 85 + 10/12 = 85.8333$), e in una data lontana dai periodi anomali elencati, ottobre 1992 ($t_0 = 92 + 10/12 = 92.8333$)⁹.

Per verificare l'effetto della data di campionamento h_1 e h_2 , valori del parametro di liscio per il calcolo dei pesi in base alla distanza delle osservazioni da x_0 , assumono valore costante ($h_1 = h_2 = 1$) mentre si fa variare h_3 ; in questo modo per commentare i grafici che seguono e analizzare le differenze delle stime al cambiare del grado di liscio imposto, si confrontano i grafici che seguono con quello riportato in Figura 3.16 (c-d).

⁹ Per calcolare i pesi in base alla distanza temporale, t e t_0 vengono espressi nella forma *anno+mese/12*.

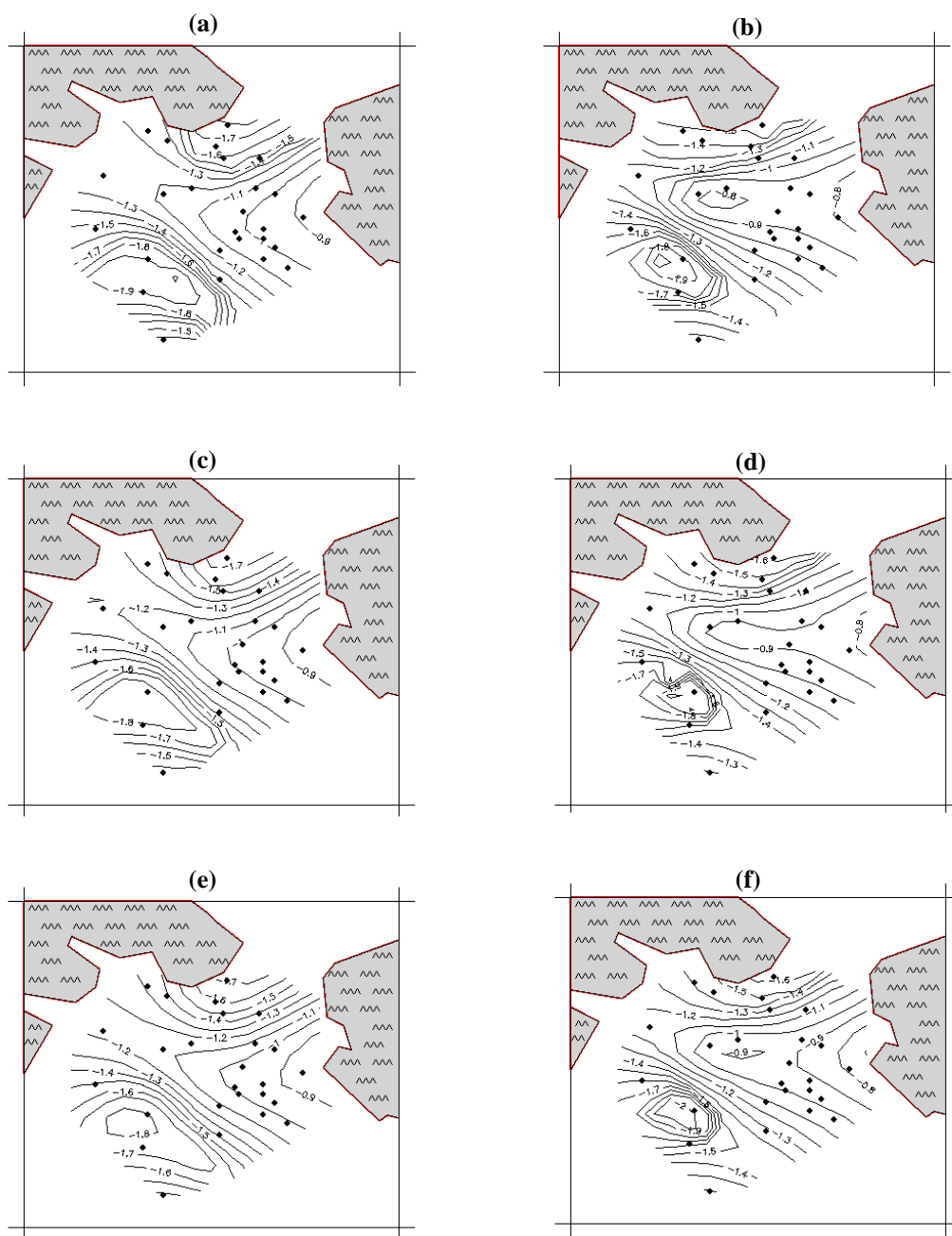


Figura 5.2 Curve di livello della stima della mediana tramite verosimiglianza locale calcolando la distanza temporale dall'ottobre 1985 (grafico di sinistra), e dall'ottobre 1992 (grafico di destra), fissando $h_3=0.2$ (a-b), $h_3=0.4$ (c-d), $h_3=0.6$ (e-f).

CONCLUSIONI

In questo progetto sono state analizzate misurazioni su campioni d'acqua prelevati da pozzi dell'isola di Vulcano. È stato considerato il log-rapporto di due componenti principali dell'acqua, il calcio e il solfato. Non avendo a disposizione elementi informativi che consentano di formulare un modello statistico di riferimento per rappresentare in maniera semplificata il rapporto $\ln((Ca^{2+})/(SO_4^{2-}))$ in prima istanza è stata effettuata un'analisi preliminare grafica. Da questa è emerso che la densità dei dati di ogni pozzo ha forma a campana, che richiama la distribuzione normale, ma con la presenza di asimmetria. Una distribuzione che soddisfa tali requisiti è la normale asimmetrica. Dopo avere confermato che tale distribuzione è adeguata per spiegare il fenomeno d'interesse, tramite il meccanismo della regressione locale è emerso che sia il parametro che regola la forma che i valori assunti dalla variabile sono fattori discriminanti per i pozzi: note le coordinate del punto si riesce ad identificare il segno dell'asimmetria e la mediana dei valori assunti dalla variabile. Tramite il metodo della verosimiglianza locale, invece, si è verificato come l'aggiunta di pesi alla log-verosimiglianza modifichi le stime dei parametri della distribuzione in base alla distanza dei punti osservati e tenga conto della loro fittezza.

Oltre alla normale asimmetrica anche la distribuzione t asimmetrica risulta adeguata a soddisfare i requisiti emersi durante l'analisi preliminare. Stimando i parametri di tale distribuzione emerge un numero elevato di pozzi i cui gradi di libertà permettono l'approssimazione alla normale asimmetrica. Approfondendo questo aspetto tramite la regressione locale è emerso che tale parametro è fattore

discriminante per individuare la distribuzione che segue meglio i dati (note le coordinate del pozzo è possibile individuare se il parametro dei gradi di libertà della relativa distribuzione assume valore sufficientemente alto per poter effettuare l'approssimazione alla normale asimmetrica).

Anche in questo caso, tramite il metodo della verosimiglianza locale, si è verificato come l'aggiunta di pesi alla log-verosimiglianza modifichi le stime dei parametri della distribuzione in base alla distanza dei punti osservati.

Si può così concludere che la distanza dei pozzi dal vulcano influisce sul rapporto tra catione calcio e anione solfato: la localizzazione di ognuno di essi permette di determinare la distribuzione che meglio segue l'andamento dei dati, il segno del parametro di asimmetria e i valori che la variabile analizzata assume.

BIBLIOGRAFIA

- Azzalini, A. e Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution, *J. R. Statist. Soc. B*, v. 61, n. 3, p.579-602.
- Azzalini, A. e Capitanio, A. (2002). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution, *J. R. Statist. Soc. B*, v. 65, n. 2, p.367-389.
- Azzalini, A. e Scarpa, B. (2004). *Analisi dei dati e data mining*, Springer-Verlag, Milano.
- Azzalini, A. e Dalla Valle, A. (1996). The multivariate skew normal distribution, *Biometrika*, v. 83, n. 4, p. 715-726.
- Bowman, A.W. e Azzalini, A. (1997). *Applied smoothing techniques for data analysis: the Kernel approach with S-Plus Illustrations*, Oxford University Press, Oxford.
- Bowman, A.W. e Azzalini, A. (2003). Computational aspects of nonparametric smoothing with illustrations from the sm library, *Computational Statistics & Data Analysis*, v. 42, p. 545-560.
- Bortot, P., Ventura, L. e Salvan, A. (2000). *Inferenza statistica: applicazioni con S-Plus e R*, Cedam, Padova.
- Buccianti, A. e Pawlowsky-Glahn, V. (2005). New perspectives on water chemistry and compositional data analysis, *Mathematical Geology*, v. 37, n. 7, p. 703-727.
- Buccianti, A. e Pawlowsky-Glahn, V. (2006). Statistical evaluation of compositional changes in volcanic gas chemistry: a case study, *Stoch. Environ. Res. Risk. Assess*, v. 21, p. 25-33.

Siti web consultati in Internet:

<http://azzalini.stat.unipd.it/SN/Intro/intro.html>
http://www.anisn.it/geologia2000/A_vulc.html
<http://www.isoleolie.org/Vulcano/Geologia.htm#logo>